

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compen Ltd.

OM protein - protein search, using sw model

Run on: July 19, 2001, 16:07:51 ; Search time 36.24 Seconds  
(without alignments)  
4472.231 Million cell updates/sec

Title: US-09-501-171A-4

Perfect score: 1225  
Sequence: 1 MFARKPGAPLGAAMPVDPQ.....PYSELNYETSHYSPDSWSV 1225

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 425026 seqs, 132305027 residues

Wor size: 0

To, number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

1: SP\_ARCHAEA:\*  
2: SP\_BACTERIA:\*  
3: SP\_FUNGI:\*  
4: SP\_HUMAN:\*  
5: SP\_INVERTEBRATE:\*  
6: SP\_MAMMAL:\*  
7: SP\_MHC:\*  
8: SP\_ORGANELLE:\*  
9: SP\_PHAGE:\*  
10: SP\_PLANT:\*  
11: SP\_RODENT:\*  
12: SP\_UNCLASSIFIED:\*  
13: SP\_VERTEBRATE:\*  
14: SP\_VIRUS:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1225	100.0	1225	4 Q9UBQ3	Q9UBQ3 homo sapien
2	868	70.9	1040	4 Q9UM66	Q9UM66 homo sapien
3	775	63.3	876	4 Q43840	Q43840 homo sapien
4	720	58.8	1225	4 Q00379	Q00379 homo sapien
5	312	25.5	321	4 Q43206	Q43206 homo sapien
6	284	23.2	1247	11 Q35927	Q35927 mus musculu
7	264	21.6	264	11 Q35116	Q35116 rattus norv
8	206	16.8	307	4 Q15390	Q15390 homo sapien
9	114	9.3	114	4 Q9UPM3	Q9UPM3 homo sapien
10	26	2.1	1211	4 Q99569	Q99569 homo sapien
11	16	1.3	779	5 Q9W579	Q9W579 drosophila
12	16	1.3	781	5 Q9NHP1	Q9NHP1 drosophila
13	15	1.2	91	4 Q95645	Q95645 homo sapien
14	11	0.9	616	4 Q9UP73	Q9UP73 homo sapien
15	11	0.9	742	13 Q9DE61	Q9DE61 xenopus lae
16	11	0.9	838	4 Q9UP71	Q9UP71 homo sapien
17	11	0.9	885	4 Q9UP72	Q9UP72 xenopus lae
18	11	0.9	907	13 Q9DFB2	Q9DFB2 xenopus lae
19	11	0.9	933	4 Q60935	Q60935 homo sapien

## ALIGNMENTS

20	11	0.9	939	4 Q60714	Q60714 homo sapien
21	11	0.9	939	4 Q15088	Q15088 homo sapien
22	11	0.9	941	4 Q60713	Q60713 homo sapien
23	11	0.9	962	4 Q60715	Q60715 homo sapien
24	11	0.9	968	4 Q60716	Q60716 homo sapien
25	11	0.9	1181	5 Q9U308	Q9U308 caenorhabdi
26	10	0.8	607	10 Q9U308	Q9U308 caenorhabdi
27	10	0.8	1336	10 Q81074	Q81074 arabidopsis
28	9	0.7	53	4 Q92728	Q92728 homo sapien
29	9	0.7	83	5 Q9GT19	Q9GT19 heterodera
30	9	0.7	141	14 Q9DUC8	Q9DUC8 lt virus. o
31	9	0.7	153	5 Q9NA70	Q9NA70 caenorhabdi
32	9	0.7	239	2 Q66118	Q66118 zymomonas m
33	9	0.7	301	10 Q9LX93	Q9LX93 arabidopsis
34	9	0.7	368	5 Q45198	Q45198 caenorhabdi
35	9	0.7	391	2 Q9RPX5	Q9RPX5 bruceia su
36	9	0.7	400	10 Q9ZRM3	Q9ZRM3 antilrhinum
37	9	0.7	409	5 Q9U0Z7	Q9U0Z7 leishmania
38	9	0.7	415	6 Q29015	Q29015 sus scrofa
39	9	0.7	443	5 Q9XYA8	Q9XYA8 dictyosteli
40	9	0.7	446	4 Q15881	Q15881 homo sapien
41	9	0.7	451	11 Q62775	Q62775 rattus norv
42	9	0.7	477	6 Q97600	Q97600 oryctolagus
43	9	0.7	479	11 Q9JMK4	Q9JMK4 mus musculu
44	9	0.7	485	11 Q9Z0G8	Q9Z0G8 rattus norv
45	9	0.7	496	4 Q9UDY7	Q9UDY7 homo sapien

RESULT	1	PRELIMINARY:	PRT: 1225 AA.
ID	Q9UBQ3		
AC	Q9UBQ3;		
DT	01-MAY-2000 (TREMBLrel. 13, Created)		
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)		
DT	01-MAR-2001 (TREMBLrel. 16, Last annotation update)		
DE	NEURAL PLAKOPHILIN-RELATED ARM-REPEAT PROTEIN (NPRAP).		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=BRAIN;		
RA	MEDLINE=99223289; PubMed=10208590;		
TX	Tanahashi H.;		
RT	Isolation of human delta-catenin and its binding specificity with		
RL	presentin 1. (1563-568(1999)).		
DR	EMBL: AB013805; BAA36187-17.		
DR	InterPro: IPR000225; .		
DR	Pfam: PF00514; Armadillo_seg. 6.		
DR	PROSITE: PS00176; ARM_REPEAT; 3.		
DR	SMART: SM00185; ARM; 1		
DR	SEQUENCE 1225 AA; 13265 MW; 4A24ACG3E22BFE83 CRC64;		

2-25-1999  
post purity check

Query Match	100.0%; Score 1225; DB 4; Length 1225;
Best Local Similarity	100.0%; Pred. No. 0;
Matches 1225;	Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 MFARKPGAPLGAAMPVDPQSSASEKTSLSFGLTSGDSEETTTAIIASVKEDEL 60
DB	1 MFARKPGAPLGAAMPVDPQSSASEKTSLSFGLTSGDSEETTTAIIASVKEDEL 60
QY	61 QFERLTFELAEARQIVASQERCKLSFGSGMSMSASAEQFOWOSQDQKIDLELTG 120
DB	61 QFERLTFELAEARQIVASQERCKLSFGSGMSMSASAEQFOWOSQDQKIDLELTG 120
QY	121 LEIVDSCTIRSLQSGILDPDYSTGERPSILSOSALQNSKPEGSFOYPASYHSGTIAL 180
DB	121 LEIVDSCTIRSLQSGILDPDYSTGERPSILSOSALQNSKPEGSFOYPASYHSGTIAL 180

```

Db 121 LELVDSICRISLQESGILUDPDYSTGERPSILSQSALQJLNSKPEGSFOYPASYSNQTIAL 180
Qy 181 GETTSPQPARQTOARATGOSTSOTTSRAGHLAGPEAPPPPPPPPPPPPPPPPPPPPPPPPP 240
Db 181 GETTSPQPARQTOARATGOSTSOTTSRAGHLAGPEAPPPPPPPPPPPPPPPPPPPPPPPPP 240
Qy 241 PDAPAAAAAALYSSSTLPAPPRGSPPLAAPOGSPKTLORGSGAPGATYAAPRGSSP 300
Db 241 PDAPAAAAAALYSSSTLPAPPRGSPPLAAPOGSPKTLORGSGAPGATYAAPRGSSP 300
Qy 301 KOSPERLAKSISTSPINIVYSSAGLSPIRYTSPTVOSTISSPITHOLSTICTYATLS 360
Db 301 KOSPERLAKSISTSPINIVYSSAGLSPIRYTSPTVOSTISSPITHOLSTICTYATLS 360
Qy 361 PTKRLVHASSEYKSHOSOLYATATLORPGSLAAGSRASYSQHGHLGELALOSPEHHT 420
Db 361 PTKRLVHASSEYKSHOSOLYATATLORPGSLAAGSRASYSQHGHLGELALOSPEHHT 420
Qy 421 DEYEDRYOKRPMKSLISQSGDPLPAHNTGTYRTSTAPSSPYDVSPLOKRTGSOHGPON 480
Db 421 DEYEDRYOKRPMKSLISQSGDPLPAHNTGTYRTSTAPSSPYDVSPLOKRTGSOHGPON 480
Qy 481 AAAAFQOASVYAGPASYADPYRQLOQCPSEVSPYSKSGPALPEGTARSPSIDSIQK 540
Db 481 AAAAFQOASVYAGPASYADPYRQLOQCPSEVSPYSKSGPALPEGTARSPSIDSIQK 540
Qy 541 DPREGMDPELPEYIOMLOHQFPVSQSNAAAYLOHLCFGDNKTKAETIRROGGIQLLYDL 600
Db 541 DPREGMDPELPEYIOMLOHQFPVSQSNAAAYLOHLCFGDNKTKAETIRROGGIQLLYDL 600
Qy 601 LDHRTVEVHSACGALNLYGKANDNKTALKNCGGIPALVRLRKTDTLEIRELVTV 660
Db 601 LDHRTVEVHSACGALNLYGKANDNKTALKNCGGIPALVRLRKTDTLEIRELVTV 660
Qy 661 LWNSSCALMPITIOALAVLTNAVLIIPHSQWNSPLQDRKTOLSSQVLRNATGCLR 720
Db 661 LWNSSCALMPITIOALAVLTNAVLIIPHSQWNSPLQDRKTOLSSQVLRNATGCLR 720
Qy 721 NVSSAGEBARRECDGLDALLVYIOSALGSSEIDSKTEVENCVCILRLNRYLAETS 780
Db 721 NVSSAGEBARRECDGLDALLVYIOSALGSSEIDSKTEVENCVCILRLNRYLAETS 780
Qy 781 OGHMGDELGLGCGANGKDASSGCKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 840
Db 781 OGHMGDELGLGCGANGKDASSGCKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 840
Qy 841 LMRHSIVKPYLTLLSECSNPDTLEGAGALQNLAAQSMKMSVYIRAAVREKKGPIIYEL 900
Db 841 LMRHSIVKPYLTLLSECSNPDTLEGAGALQNLAAQSMKMSVYIRAAVREKKGPIIYEL 900
Qy 901 LRINDRYVCAVATALRNALDVANKELIGYAMRDLYHRLPGGNNNSNTASKAMSDTY 960
Db 901 LRINDRYVCAVATALRNALDVANKELIGYAMRDLYHRLPGGNNNSNTASKAMSDTY 960
Qy 961 TAVCTTJHEVITKMEKAKALRDAGIEKLVGISKSGDKHSPKYVAAQOVLTNSMQR 1020
Db 961 TAVCTTJHEVITKMEKAKALRDAGIEKLVGISKSGDKHSPKYVAAQOVLTNSMQR 1020
Qy 1021 DLRSVKKKDSOYHFAVASSSTTERORPYSSSRTPSISPVAVSPNNRSASAPSPREX 1080
Db 1021 DLRSVKKKDSOYHFAVASSSTTERORPYSSSRTPSISPVAVSPNNRSASAPSPREX 1080
Qy 1081 ISLKERKTEVETGNSNATYGAKEHTSRKDAVTAQNTGISTLYRNSYGAPADIKHNOY 1140
Db 1081 ISLKERKTEVETGNSNATYGAKEHTSRKDAVTAQNTGISTLYRNSYGAPADIKHNOY 1140
Qy 1141 SAQVPEPESKRYETQOPONSTRNNTDESEFEDOVHHRPASEYMHGKSTGVYDF 1200
Db 1141 SAQVPEPESKRYETQOPONSTRNNTDESEFEDOVHHRPASEYMHGKSTGVYDF 1200
Qy 1201 YSAAAPYSSELYNETSHYPASPSDMV 1225
Db 1201 YSAAAPYSSELYNETSHYPASPSDMV 1225

```

```

RESULT 2
ID Q9DM66 PRELIMINARY; PRT: 1040 AA.
AC Q9UM66;
DT 01-MAY-2000 (TREMBLrel. 13' Created)
DT 01-MAY-2000 (TREMBLrel. 13', last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, last annotation update)
DE G724 (FRAGMENT).
GN G724.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OC NCBI_TaxId=9606;
RN 11)
RP SEQUENCE FROM N.A.
RA Levesque G., Yu G., Fraser P.E., StGeorge-Hyslop P.;
RT "Presentin 1 interacts with a novel protein that contains armadillo
RT repeats and maps near the Cri du chat locus on chromosome 5p.";
RL Submitted (DEC-1996) to the EMBL/Genbank/DBJ databases.
DR EMBL: U61004; AAD00453.1; -.
DR InterPro: IPR000225; -.
DR Pfam: PF00514; Armadillo_seg; 6.
DR PROSITE: PS00176; ARM_REPEAT; 3.
DR SMART: SM00185; ARM; 1.
DR NON_TER 1.
SQ SEQUENCE 1040 AA; 112802 MW; E12C8694EAC51D42 CRC64;

```

```

Query Match 70.9%; Score 868; DB 4; Length 1040;
Best local similarity 99.9%; Pred. No. 0;
Matches 968; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

Qy 257 STLPAPRGSPPLAAPOGSPKTLORGSGAPGATYAAPRGSSPKQSPRLAKSYSTSP 316
Db 72 STLPAPRGSPPLAAPOGSPKTLORGSGAPGATYAAPRGSSPKQSPRLAKSYSTSP 131
Qy 317 INIVYSSAGLSPIRYTSPTVOSTISSPITHOLSTICTYATLSPTKRLVHASSEYKSH 376
Db 132 INIVYSSAGLSPIRYTSPTVOSTISSPITHOLSTICTYATLSPTKRLVHASSEYKSH 191
Qy 377 QELYATATLORPGSLAAGSRASYSQHGHLGELALOSPEHHTDIYEDRYOKRPMKRS 436
Db 192 QELYATATLORPGSLAAGSRASYSQHGHLGELALOSPEHHTDIYEDRYOKRPMKRS 251
Qy 437 LSQSGDPLPAHNTGTYRTSTAPSSPYDVSPLOKRTGSOHGPONNAAATFQOASVYAGPA 496
Db 252 LSQSGDPLPAHNTGTYRTSTAPSSPYDVSPLOKRTGSOHGPONNAAATFQOASVYAGPA 311
Qy 497 SNYADPYRQLOQCPSEVSPYSKSGPALPEGTARSPSIDSIQKDPREFGWRDPELPEVI 556
Db 312 SNYADPYRQLOQCPSEVSPYSKSGPALPEGTARSPSIDSIQKDPREFGWRDPELPEVI 371
Qy 557 QMLQHOFPVSQSNAAAYLOHLCFGDNKTKAETIRROGGIQLLYDLIDHRTVEVHSACAL 616
Db 312 QMLQHOFPVSQSNAAAYLOHLCFGDNKTKAETIRROGGIQLLYDLIDHRTVEVHSACAL 431
Qy 617 RULVYKANDNKTALKNCGGIPALVRLRKTDTLEIRELVTVGLWLNSSCDALKMPITIO 676
Db 432 RULVYKANDNKTALKNCGGIPALVRLRKTDTLEIRELVTVGLWLNSSCDALKMPITIO 491
Qy 677 DALAVLTNAVLIIPHSQWNSPLQDRKTOLSSQVLRNATGCLNNVSSAGEBARRRREC 736
Db 492 DALAVLTNAVLIIPHSQWNSPLQDRKTOLSSQVLRNATGCLNNVSSAGEBARRRREC 551
Qy 737 DGLDALLVYIOSALGSSEIDSKTEVENCVCILRLNRYLAETSQGHMGTDDELGLGCG 796
Db 552 DGLDALLVYIOSALGSSEIDSKTEVENCVCILRLNRYLAETSQGHMGTDDELGLGCG 611
Qy 797 EANGKDASSGCKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 856
Db 612 EANGKDASSGCKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 671

```

```

QY 857 CSNPDTLEGAAGALONTLAAGSMKMSVYIRAAVRKEKGLPIVLELLRIDNDRVCAVATL 916
DB 672 CSNPDTLEGAAGALONTLAAGSMKMSVYIRAAVRKEKGLPIVLELLRIDNDRVCAVATL 731
QY 917 RNNALDVNRKELIGKAYMBDLVHRLPGNNSNNTASKANSDDTVAVCTTLEHVTTKNNE 976
DB 732 RNNALDVNRKELIGKAYMBDLVHRLPGNNSNNTASKANSDDTVAVCTTLEHVTTKNNE 791
QY 977 NAKALDAGGIEKLVGISKSGDKHSPKVVKAASQVLYNSMMQYRDLRSYKKDGMSQYHF 1036
DB 792 NAKALDAGGIEKLVGISKSGDKHSPKVVKAASQVLYNSMMQYRDLRSYKKDGMSQYHF 851
QY 1037 VASSTTERDROPRYSSTSPISPRVSPNNRSASAPSPREMI SLKRRKTDYECTGSN 1096
DB 852 VASSTTERDROPRYSSTSPISPRVSPNNRSASAPSPREMI SLKRRKTDYECTGSN 911
QY 1097 ATYHGAKEHTSRKDMTAONTGISTLYNSYGAPADEDIKHNOVSAQVPQEPSRKDYET 1156
DB 912 ATYHGAKEHTSRKDMTAONTGISTLYNSYGAPADEDIKHNOVSAQVPQEPSRKDYET 971
QY 957 YQFPONSTRNYDSFFEDQVHHRPPASEYTMHLGLKSTGNYDVFYSAAAPYSELNYESH 1216
DB 972 YQFPONSTRNYDSFFEDQVHHRPPASEYTMHLGLKSTGNYDVFYSAAAPYSELNYESH 1031
QY 1217 YPASPSDW 1225
DB 1032 YPASPSDW 1040

RESULT 3
043840 PRELIMINARY: PRT: 876 AA.
AC 043840;
DT 01-JUN-1998 (Tremblrel. 06, Created)
DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE NEURAL PLAKOPHTLIN RELATED ARM-REPEAT PROTEIN (FRAGMENT).
GN NRPAP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=98002299; PubMed=9342840;
RA Paffenholz R., Franke W.W.;
RT "Identification and localization of a neurally expressed member of the
RT Lakoglobin/armadillo multigene family.";
RL Differentiation 61:293-304(1997).
DR EMBL: U52351; AAB97957.1; -.
DR InterPro: IPR000225; -.
DR Pfam: PF00514; Armadillo_seg. 6.
DR PROSITE: PSS0176; ARM_REPEAT. 3.
DR SMART: SM00185; ARM. 1.
FT NON_CODING 1
FT SEQUENCE 876 AA; 96442 MW; 8D2342B71CC9E92A CRC64;
SQ

```

Query Match 63.3%; Score 775; DB 4; Length 876;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 875; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

DB 121 QRTSOHQHPQNAATAATFORASYAAGPASNVADPYROLQCPVSEPSYKSGPALPREGTL 180
QY 530 ARSPSIDSIQDPRFEGMRPELPEVIOMLQHOPFVSQNAAYLOHICGDNKIKAEIR 589
DB 181 ARSPSIDSIQDPRFEGMRPELPEVIOMLQHOPFVSQNAAYLOHICGDNKIKAEIR 240
QY 590 ROGGIQLLVLDLDRHMTVEHRSACGALRNLYGKANDDNKIALKNCGGITAYRLKRT 649
DB 241 ROGGIQLLVLDLDRHMTVEHRSACGALRNLYGKANDDNKIALKNCGGITAYRLKRT 300
QY 650 DLEIRELVTVGLNMLSSCDALKMPITODALAVLTNAVITPHSGENSLDODRKTOLHSS 709
DB 301 DLEIRELVTVGLNMLSSCDALKMPITODALAVLTNAVITPHSGENSLDODRKTOLHSS 360
QY 710 QVLRNATGCLRNVSAGGEARRRRECGLTDALLIYIOSALGSESEIDSKTVENCYILR 769
DB 361 QVLRNATGCLRNVSAGGEARRRRECGLTDALLIYIOSALGSESEIDSKTVENCYILR 420
QY 770 NLSYRLAAETSGOHMGTDLDGLGCGANGKDAESSGCGWKKKKKKSODQMDGVPLP 829
DB 421 NLSYRLAAETSGOHMGTDLDGLGCGANGKDAESSGCGWKKKKKKSODQMDGVPLP 480
QY 830 DCAEPPKGIOMLMPISYKPYTLTISECSNPDTLEGAAGALONTLAAGSMKMSVYIRAAVR 889
DB 481 DCAEPPKGIOMLMPISYKPYTLTISECSNPDTLEGAAGALONTLAAGSMKMSVYIRAAVR 540
QY 890 KEKGLPIVLELLRIDNDRVCAVATLARNMALDYRNKELIKYAMRDVHRLPGNNSNN 949
DB 541 KEKGLPIVLELLRIDNDRVCAVATLARNMALDYRNKELIKYAMRDVHRLPGNNSNN 600
QY 950 TASKAMSDDTVAVCTTLEHVTTKMENAKALRDAGIEKLVGISKSGDKHSPKVVKA 1009
DB 601 TASKAMSDDTVAVCTTLEHVTTKMENAKALRDAGIEKLVGISKSGDKHSPKVVKA 660
QY 1010 SOVLNSMMQYRDLRSYKKDGMSQYHFVASSSTTERDQORYSSRTPSPVAVSPNNR 1069
DB 661 SOVLNSMMQYRDLRSYKKDGMSQYHFVASSSTTERDQORYSSRTPSPVAVSPNNR 720
QY 1070 SASAPSPREMI SLKERTDYECTGSNATYHGAKEHTSRKDMTAONTGISTLYNSYG 1129
DB 721 SASAPSPREMI SLKERTDYECTGSNATYHGAKEHTSRKDMTAONTGISTLYNSYG 780
QY 1130 APAEDIKHNOVSAQVPQEPSRKDYETQFPONSTRNYDESFFEDQVHHRPPASEYTMHL 1189
DB 781 APAEDIKHNOVSAQVPQEPSRKDYETQFPONSTRNYDESFFEDQVHHRPPASEYTMHL 840
QY 1190 GLKSTGNYDVFYSAAAPYSELNYESHYPASPSDW 1225
DB 841 GLKSTGNYDVFYSAAAPYSELNYESHYPASPSDW 876

RESULT 4
000379 PRELIMINARY: PRT: 1225 AA.
AC 000379; Q13589;
DT 01-JUL-1997 (Tremblrel. 04, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE DELTA-CATENIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=97366296; PubMed=9223106;
RA Zhou J., Liyanage U., Medina M., Ho C., Simmons A.D., Kosik K.S.;
RT "Presentin 1 interaction in the brain with a novel member of the
RT Armadillo family.";
RL Neuroreport 8:2085-2090(1997).
RN [2]
RP SEQUENCE FROM N.A.

```

RC TISSUE-BRAIN;  
 RA Lu Q., Paredes M., Medina M., Zhou J., Cavallo R., Peifer M.,  
 RA Orecchio L., Kosik K.S.;  
 RT "d-catenin, an adhesive junction associated protein which promotes  
 RT motile behavior";  
 RL Submitted (Oct-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 865-1225 FROM N.A.  
 RA MEDLINE=97202103; PubMed=9049630;  
 RA Simmons A.D., Overhauser J., Lovett M.;  
 RT "Isolation of cDNAs from the Cri-du-chat critical region by direct  
 RT screening of a chromosome 5-specific cDNA library."  
 RL Genome Res. 7:118-127(1997).  
 DR EMBL; U96136; AAC63103.1; -;  
 DR EMBL; U52828; AAB96357.1; -;  
 DR InterPro: IPR000225; -;  
 DR Pfam: PF00514; Armadillo\_seg. 6.  
 DR PROSITE: PS50176; ARM\_REPEAT; 3.  
 DR SMART: SM00185; ARM; 1.  
 SQ SEQUENCE 1225 AA; 132664 MW; 8B676CFD5AFA4E65 CRC64;

Query Match 58.8%; Score 720; DB 4; Length 1225;

Best Local Similarity 99.6%; Pred. NO. 0;  
 Matches 1220; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 MFARKPGAGAPGAPYDPDPSSASEKTSLSPLGINTSGDSETEYTSAILASVKEOEL 60  
 Db 1 MFARKPGAGAPGAPYDPDPSSASEKTSLSPLGINTSGDSETEYTSAILASVKEOEL 60  
 QY 61 QEBELTELEERIVASOLEKCLGSETSMSSMAEFOFOWOSODGKIDEDLTG 120  
 Db 61 QEBELTELEERIVASOLEKCLGSETSMSSMAEFOFOWOSODGKIDEDLTG 120  
 QY 121 LEIVDSICIRLSQESGILDPDYSTERPILLSOSALQINKPEGSFOYPASYNSOTIAL 180  
 Db 121 LEIVDSICIRLSQESGILDPDYSTERPILLSOSALQINKPEGSFOYPASYNSOTIAL 180  
 QY 181 GETTSOLPARGTORATGGSFSGGTSRAGHLAGEPPAPPPPPPREPPAPSLGSAFHL 240  
 Db 181 GETTSOLPARGTORATGGSFSGGTSRAGHLAGEPPAPPPPPPREPPAPSLGSAFHL 240  
 QY 241 PDAPPAALAAALYSSSTPLAPRGSSPLAAPGGSPTKIQRGSAPEGATYAAPGSSP 300  
 Db 241 PDAPPAALAAALYSSSTPLAPRGSSPLAAPGGSPTKIQRGSAPEGATYAAPGSSP 300  
 QY 301 KQSPSLAKSYSTSSPINIVSSAGLSPIRVTPPVOSTISSPIHQLSSTIGTATLS 360  
 Db 301 KQSPSLAKSYSTSSPINIVSSAGLSPIRVTPPVOSTISSPIHQLSSTIGTATLS 360  
 QY 361 PTRKLVASQYKSHSOELATATLQPGSLAAGSRAVSQGHGPELALQSPENHI 420  
 Db 361 PTRKLVASQYKSHSOELATATLQPGSLAAGSRAVSQGHGPELALQSPENHI 420  
 QY 421 DPYEDVYQKPMRSLSOSGDPPLPAHNGYRTSTAPSSPEVDVPIQRTSQGPPON 480  
 Db 421 DPYEDVYQKPMRSLSOSGDPPLPAHNGYRTSTAPSSPEVDVPIQRTSQGPPON 480  
 QY 481 AAATFORASAYAPASVADPYROLQYCPVSPEYKSGPALPREGTILARSPIDSIOK 540  
 Db 481 AAATFORASAYAPASVADPYROLQYCPVSPEYKSGPALPREGTILARSPIDSIOK 540  
 QY 541 DPREFGMRDELEVIOMLOHOPPSVQSNAAAYLOHLCFGDNKIKAEIRROGCIQLVLD 600  
 Db 541 DPREFGMRDELEVIOMLOHOPPSVQSNAAAYLOHLCFGDNKIKAEIRROGCIQLVLD 600  
 QY 601 LDRHMTFVHSAGALRNLYGKANDNKITALKNGGISTALVRLKTDLEIRELVTV 660  
 Db 601 LDRHMTFVHSAGALRNLYGKANDNKITALKNGGISTALVRLKTDLEIRELVTV 660  
 QY 661 LNMVSSCDALKMPDIIDALAVLTVNAVIPHSGWENSPLODDKRIQLHSSQVARNATGCLR 720  
 Db 661 LNMVSSCDALKMPDIIDALAVLTVNAVIPHSGWENSPLODDKRIQLHSSQVARNATGCLR 720

QY 721 NVSSAGEPARRRREDCGTLTDLALIVYOSALSSSEIDSKTYENCYCILRNSTYLAETS 780  
 Db 721 NVSSAGEPARRRREDCGTLTDLALIVYOSALSSSEIDSKTYENCYCILRNSTYLAETS 780  
 QY 781 QGQHMGTDELGLLGEANGKDAESSGCGKKKKKKKSODDMDGVPLPDCAEPKGIOM 840  
 Db 781 QGQHMGTDELGLLGEANGKDAESSGCGKKKKKKKSODDMDGVPLPDCAEPKGIOM 840  
 QY 841 LHPSTIVKPYLTLLSECSNPDTLEGAAGALQNLAAAGSMKWSYIIRAAVKEKGLPIVEL 900  
 Db 841 LHPSTIVKPYLTLLSECSNPDTLEGAAGALQNLAAAGSMKWSYIIRAAVKEKGLPIVEL 900  
 QY 901 LRTDNRVACAVATLRNALDLVRKELIGKAMADLYHRLPGGNSNNTASKMSDPTV 960  
 Db 901 LRTDNRVACAVATLRNALDLVRKELIGKAMADLYHRLPGGNSNNTASKMSDPTV 960  
 QY 961 TAVCCTLHEVITKMMENAKALRDAGIEKLVGISKSKDKHSPKVKAAQVLSMMQYR 1020  
 Db 961 TAVCCTLHEVITKMMENAKALRDAGIEKLVGISKSKDKHSPKVKAAQVLSMMQYR 1020  
 QY 1021 DLRSLYKKDGMSQYHFVASSSTIERDORPYSSSTPSISPVRSVNNRSASAPAPREM 1080  
 Db 1021 DLRSLYKKDGMSQYHFVASSSTIERDORPYSSSTPSISPVRSVNNRSASAPAPREM 1080  
 QY 1081 ISLKERKTDYECTGSNATYHGAKGHTSRKDMTQONTGISITVRSYGAPAEIDKHMV 1140  
 Db 1081 ISLKERKTDYECTGSNATYHGAKGHTSRKDMTQONTGISITVRSYGAPAEIDKHMV 1140  
 QY 1141 SAQVPQEPESRKDYETYPONSTRNRYDESEFEDVHHRPAPSEYTHMLGLKSTGVYDF 1200  
 Db 1141 SAQVPQEPESRKDYETYPONSTRNRYDESEFEDVHHRPAPSEYTHMLGLKSTGVYDF 1200  
 QY 1201 YSARPYSELNETHYPASPDW 1225  
 Db 1201 YSARPYSELNETHYPASPDW 1225

RESULT 5  
 ID 043206 PRELIMINARY; PRT; 321 AA.  
 AC 043206;  
 DT 01-JUN-1998 (TREMBLrel. 06, Created)  
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)  
 DE PROBABLE CATEININ.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-BRAIN.  
 RX MEDLINE=96207227; PubMed=8619474;  
 RA Anderson B., Wentland M.A., Ricafrente J.Y., Liu W., Gibbs R.A.;  
 RT "A 'double adaptor' method for improved shotgun library  
 RT construction.";  
 RL Anal. Biochem. 236:107-113(1996).  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-BRAIN.  
 RX MEDLINE=97264341; PubMed=9110174;  
 RA Yu W., Anderson B., Worley K.C., Muzny D.M., Ding Y., Liu W.,  
 RA Ricafrente J.Y., Wentland M.A., Lennon G., Gibbs R.A.;  
 RT "Large-scale concatenation cDNA sequencing.";  
 RL Genome Res. 7:353-358(1997).  
 DR EMBL; AF035302; AAB88185.1; -;  
 KW Cytoskeleton; Structural protein; Cell adhesion.  
 SQ SEQUENCE 321 AA; 36108 MW; 1DDF4811891DA053 CRC64;

Query Match 25.5%; Score 312; DB 4; Length 321;  
 Best Local Similarity 100.0%; Pred. No. 0;

Matches 312; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 914 TAI RNALADVRNKLIGKYMADLVRLDGGNNNTASKAMDDPVTAVCCITLHEVIRK 973  
 DB 10 TAI RNALADVRNKLIGKYMADLVRLDGGNNNTASKAMDDPVTAVCCITLHEVIRK 69  
 QY 974 NMENAKALRDAGGIEKLVGSKSGDKHSPKVVKAASQVLSNMWQYRDLRSILKKDQMSQ 1033  
 DB 70 NMENAKALRDAGGIEKLVGSKSGDKHSPKVVKAASQVLSNMWQYRDLRSILKKDQMSQ 129  
 QY 1034 YHFAVSSSTTERDORPYSSTSPISPVAVSPNNRSASAPASPREMISLKEKTDYECT 1093  
 DB 130 YHFAVSSSTTERDORPYSSTSPISPVAVSPNNRSASAPASPREMISLKEKTDYECT 189  
 QY 1094 GSNATYGAAGEHTRSKDAMTAONTGISTLRNSYGAPADIDKHNOVASAOPVQEPSRKD 1153  
 DB 190 GSNATYGAAGEHTRSKDAMTAONTGISTLRNSYGAPADIDKHNOVASAOPVQEPSRKD 249  
 QY 1154 YETVQPNSTRNVDSEFEDQVHHRPASEYTMHLGKSTGVYVDFYSARPSELNVE 1213  
 DB 250 YETVQPNSTRNVDSEFEDQVHHRPASEYTMHLGKSTGVYVDFYSARPSELNVE 309  
 QY 1214 TSHYPASPDSDMV 1225  
 DB 310 TSHYPASPDSDMV 321

RESULT 6  
 ID 035927 PRELIMINARY; PRT: 1247 AA.  
 AC 035927;  
 DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE NEURAL PLAKOPHILIN-RELATED ARM-REPEAT PROTEIN.  
 GN CTRND2 OR NEBPAP.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_Taxid=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BRAIN;  
 RX MEDLINE=98002299; Pubmed=9342840;  
 RA Patfenholz R., Franke W.W.;  
 RT "Identification and localization of a neurally expressed member of the  
 RL plakoglobin/armadillo multigene family.";  
 RL Differentiation 61:293-304(1997).  
 DR EMBL, U90331; AAB82409.1; -;  
 DR SD, MG1:1195966; Ctrnd2.  
 DR InterPro: IPR000225; -;  
 DR Pfam: PF00514; Armadillo\_seg; 6.  
 DR PROSITE: PS50176; ARM\_REPEAT; 3.  
 DR SMART: SM00185; ARM; 1.  
 SQ SEQUENCE 1247 AA; 134998 MW; DA7A6B6A27D2919 CRC64;

Query Match 23.2%; Score 284; DB 11; Length 1247;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-278;  
 Matches 284; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 514 SPYKSGPALPPEGLTARSPSIDSIQKDPREFGWRDPLPEVITOMLQHPFSVOSNAAY 573  
 DB 511 SPYKSGPALPPEGLTARSPSIDSIQKDPREFGWRDPLPEVITOMLQHPFSVOSNAAY 570  
 QY 574 LOHLCEFGNKKIAETIRROGGIQLVLDLHDHMTVEHRSAGALRNLYGKANDNKIALK 633  
 DB 571 LOHLCEFGNKKIAETIRROGGIQLVLDLHDHMTVEHRSAGALRNLYGKANDNKIALK 630  
 QY 634 NCGGIPALVRLRKTDTLEIRELVTVGLVNLSSCDALKMPIIDALAVLNAVVIIPHSGW 693  
 DB 631 NCGGIPALVRLRKTDTLEIRELVTVGLVNLSSCDALKMPIIDALAVLNAVVIIPHSGW 690

QY 694 ENSPLQDDRKIQLHSSQVLRNATGCLRNVSSAGEARRRMECDGLTDALLVYQSALGS 753  
 DB 691 ENSPLQDDRKIQLHSSQVLRNATGCLRNVSSAGEARRRMECDGLTDALLVYQSALGS 750  
 QY 754 SEIDSKTYENCVCILRNLSYRLAETSGOHMGTDLDGLLGE 797  
 DB 751 SEIDSKTYENCVCILRNLSYRLAETSGOHMGTDLDGLLGE 794

RESULT 7  
 ID 035116 PRELIMINARY; PRT: 264 AA.  
 AC 035116;  
 DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE DELTA-CATENIN (FRAGMENT).  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_Taxid=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MISTAR; TISSUE=BRAIN;  
 RA Tananashi H.;  
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL, AB008752; BAA23384.1; -;  
 DR InterPro: IPR000225; -;  
 DR Pfam: PF00514; Armadillo\_seg; 2.  
 DR PROSITE: PS50176; ARM\_REPEAT; 2.  
 DR SMART: SM00185; ARM; 1.  
 FT NON\_TER 1 264  
 FT NON\_TER 264 264  
 SQ SEQUENCE 264 AA; 28928 MW; 065A86B8BD2F58E0 CRC64;

Query Match 21.6%; Score 264; DB 11; Length 264;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-258;  
 Matches 264; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 563 FPSVQSNAAAYLQHLCEFGNKKIAETIRROGGIQLVLDLHDHMTVEHRSAGALRNLYG 622  
 DB 1 FPSVQSNAAAYLQHLCEFGNKKIAETIRROGGIQLVLDLHDHMTVEHRSAGALRNLYG 60  
 QY 623 KANDDNKIALKNCGGIPALVRLRKTDTLEIRELVTVGLVNLSSCDALKMPIIDALAVL 682  
 DB 61 KANDDNKIALKNCGGIPALVRLRKTDTLEIRELVTVGLVNLSSCDALKMPIIDALAVL 120  
 QY 683 TNVVIIPHSGWENSPLODDRKIQLHSSQVLRNATGCLRNVSSAGEARRRMECDGLTDA 742  
 DB 121 TNVVIIPHSGWENSPLODDRKIQLHSSQVLRNATGCLRNVSSAGEARRRMECDGLTDA 180  
 QY 743 LLYVIOALGSSSEIDSKTYENCVCILRNLSYRLAETSGOHMGTDLDGLLGEANGKD 802  
 DB 181 LLYVIOALGSSSEIDSKTYENCVCILRNLSYRLAETSGOHMGTDLDGLLGEANGKD 240  
 QY 803 AESSGCGKKKKKKKSODQDVG 826  
 DB 241 AESSGCGKKKKKKKSODQDVG 264

RESULT 8  
 ID 015390 PRELIMINARY; PRT: 307 AA.  
 AC 015390;  
 DT 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DT 01-AUG-1998 (TREMBLrel. 07, Last annotation update)  
 DE G724.  
 GN G724.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BRAIN;  
 RA Fraser P.E., Levesque G., Rogueva E.A., Yu G., St George-Hyslop P.H.,  
 RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U72655; AAB6859.1; -  
 SQ SEQUENCE 307 AA; 34417 MW; 35640DA6113E65F CRC64;

Query Match 16.8%; Score 206; DB 4; Length 307;  
 Best Local Similarity 99.7%; Pred. No. 7.8e-200;  
 Matches 306; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 919 MALDVRNKLKLYKAMRLVHRLPGNNSNNTASKAMSDDTVTAVCTLHEVITKNNENA 978  
 DB 1 MALDVRNKLKLYKAMRLVHRLPGNNSNNTASKAMSDDTVTAVCTLHEVITKNNENA 60  
 QY 979 KALRDAGGIEKLVIGISKSGDKHSPKVVKAASQVLSNMWQYRDLRLSLYKKDQMSQYHFA 1038  
 DB 61 KALRDAGGIEKLVIGISKSGDKHSPKVVKAASQVLSNMWQYRDLRLSLYKKDQMSQYHFA 120  
 QY 1039 SSSITERORORPYSSTPSPISPVVSPNNRSASAPSPREMIISLKEKTDYECTGSNAT 1098  
 DB 121 SSSITERORORPYSSTPSPISPVVSPNNRSASAPSPREMIISLKEKTDYECTGSNAT 180  
 QY 1099 YHGAGERTSRKDMTAONTGISTLYRNSYGAPADIKHNOVSAOPVQPEPSRKDYENTQ 1158  
 DB 181 YHGAGERTSRKDMTAONTGISTLYRNSYGAPADIKHNOVSAOPVQPEPSRKDYENTQ 240  
 QY 1159 PFONSTRNYDSSFEDQVHRRPASEYTMHLGLKSTGVYDFYSANRYSSELYSHYP 1218  
 DB 241 PFONSTRNYDSSFEDQVHRRPASEYTMHLGLKSTGVYDFYSANRYSSELYSHYP 300  
 QY 1219 ASPDSMW 1225  
 DB 301 ASPDSMW 307

RESULT 9  
 Q9UPM3 PRELIMINARY; PRT; 114 AA.  
 AC Q9UPM3;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE DELTA-CATENIN (FRAGMENT).  
 OS Homo sapiens (Human).  
 OC Chkaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA Wilson, R., Smith, A., Elliott, G., Kramer, J., Latreille, P., Keppeler D.:  
 RT "The sequence of H. sapiens BAC clone RG180P08A.";  
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Waterston R.;  
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC003089; AAB83940.1; -  
 DR InterPro: IPR000225; -  
 DR Pfam: PF00514; Armadillo\_seg. 3.  
 DR PROSITE; PS50176; ARM\_REPEAT; 2.  
 FT NON\_TER 1  
 FT TER 1  
 SQ SEQUENCE 114 AA; 12961 MW; 66008C00A178D35D CRC64;

Query Match 9.3%; Score 114; DB 4; Length 114;  
 Best Local Similarity 100.0%; Pred. No. 5.9e-107;  
 Matches 114; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 543 REFGRDPELPEVYIOMLOHOFPSVQSNAAVYLOHLCFGDNKTKAETRRGGIOLVLDLD 602  
 DB 1 REFGRDPELPEVYIOMLOHOFPSVQSNAAVYLOHLCFGDNKTKAETRRGGIOLVLDLD 60  
 QY 603 HRMTVEHRSACGLRLNLYGKANDDKIKLKNCGIPALVRLRTTDEIEL 656  
 DB 61 HRMTVEHRSACGLRLNLYGKANDDKIKLKNCGIPALVRLRTTDEIEL 114

RESULT 10  
 Q99569 PRELIMINARY; PRT; 1211 AA.  
 AC Q99569;  
 DT 01-MAY-1997 (TREMBLrel. 03, Created)  
 DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE P0071 PROTEIN.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=FRONTAL CORTEX;  
 RX MEDLINE=97092329; PubMed=8937994;  
 RA Hatfield M., Nachtsheim C.;  
 RT "Cloning and characterization of a new armadillo family member, p0071,  
 RT associated with the junctional plaque: evidence for a subfamily of  
 RT closely related proteins.";  
 RL J. Cell Sci. 109:2767-2778(1996).  
 DR EMBL; X81889; CAA57478.1; -  
 DR InterPro: IPR000225; -  
 DR Pfam: PF00514; Armadillo\_seg. 7.  
 DR PROSITE; PS50176; ARM\_REPEAT; 3.  
 DR SMART; SM00185; ARM; 1.  
 SQ SEQUENCE 1211 AA; 134268 MW; 14C962691A47024F CRC64;

Query Match 2.1%; Score 26; DB 4; Length 1211;  
 Best Local Similarity 100.0%; Pred. No. 6.7e-17;  
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 913 ATALRMALDVNRKELIGYKAMRDLY 938  
 DB 893 ATALRMALDVNRKELIGYKAMRDLY 918

RESULT 11  
 Q9W5T9 PRELIMINARY; PRT; 779 AA.  
 AC Q9W5T9;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)  
 DE PUTATIVE P120CTN PROTEIN (CG17484 PROTEIN).  
 GN P120CTN OR CG17484.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BERKELEY;  
 RX MEDLINE=20196006; PubMed=10731137;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abtill J.F., Agbayani A., An H.-J., Andrews-Plannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu I., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borikova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,  
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Foster C., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,  
 RA Gaidzik A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,  
 RA Jajali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Mlshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacle J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Rhue B.C., Spadling A.C., Stapleton M., Skupski M.P., Smith T.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Chibb R.A., Myers E.W., Rubin G.M., Venter J.C.,  
 RT "The genome sequence of *Drosophila melanogaster*."  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE002751; AAF5461.1; -  
 DR HSSP: Q02248; 2BCT.  
 DR FLYBase: FBgn0015587; p120ctn.  
 DR InterPro: IPR000225; -  
 DR Pfam: PF00514; Armadillo\_seg; 4.  
 DR PROSITE: PSS0176; ARM\_REPEAT; 3.  
 DR SMART: SM00185; ARM; 1.  
 DR K W Hypothetical protein.  
 SQ SEQUENCE 779 AA; 86588 MW; 5A1F546AE055CC2B CRC64;

Query Match 1.3%; Score 16; DB 5; Length 779;  
 Best Local Similarity 100.0%; Pred. No. 6.3e-07;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 759 KTVENCVCILRNLSYR 774  
 DB 397 KTVENCVCILRNLSYR 412

RESULT 12  
 ID 09NHP1 PRELIMINARY; PRT; 781 AA.  
 AC 09NHP1;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
 DE ARM REPEAT PROTEIN.  
 GN P120CTN OR CG17484.  
 OS *Drosophila melanogaster* (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydriidae; Drosophilidae; Drosophila.  
 NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Cavallo R., Myster S., Peifer M.:  
 RT "A *Drosophila melanogaster* homolog of the adherens junction protein  
 p120ctn."  
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF220496; AAF33245.1; -  
 DR FLYBase: FBgn0015587; p120ctn.  
 DR InterPro: IPR000225; -

DR Pfam: PF00514; Armadillo\_seg; 5.  
 DR PROSITE: PSS0176; ARM\_REPEAT; 3.  
 DR SMART: SM00185; ARM; 1.  
 SQ SEQUENCE 781 AA; 86836 MW; 5F8888EA349C7CEE CRC64;

Query Match 1.3%; Score 16; DB 5; Length 781;  
 Best Local Similarity 100.0%; Pred. No. 6.3e-07;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 759 KTVENCVCILRNLSYR 774  
 DB 429 KTVENCVCILRNLSYR 444

RESULT 13  
 ID 095645 PRELIMINARY; PRT; 91 AA.  
 AC 095645;  
 DT 01-MAY-1999 (TREMBlrel. 10, Created)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)  
 DT 01-MAY-1999 (TREMBlrel. 10, Last annotation update)  
 DE PSILY2H-25 PROTEIN (FRAGMENT).  
 GN PSILY2H-25.  
 OS *Homo sapiens* (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Lavesque G., Yu G., Fraser P.E., StGeorge-Hyslop P.:  
 RT "Presentin 1 interacts with a novel protein which contains armadillo  
 RT repeats and maps near the Cri du chat locus on chromosome 9p."  
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U81005; AAD00454.1; -  
 FT NON\_TER  
 SQ SEQUENCE 91 AA; 10606 MW; 06045C43A5CD807C CRC64;

Query Match 1.2%; Score 15; DB 4; Length 91;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-06;  
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 569 NAAAYLQHLCEGDNK 583  
 DB 34 NAAAYLQHLCEGDNK 48

RESULT 14  
 ID 09UP73 PRELIMINARY; PRT; 616 AA.  
 AC 09UP73;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-MAR-2001 (TREMBlrel. 16, Last annotation update)  
 DE P120 CATENIN ISOFORM 4AC.  
 GN CTNND1.  
 OS *Homo sapiens* (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=96317528; PubMed=9653641;  
 RA Kellerschick A., Bonne S., Staes K., van Hengel J., Nollet F.,  
 RA Reynolds A., van Roy F.:  
 RT "Molecular cloning of the human p120ctn catenin gene (CTNND1):  
 RT expression of multiple alternatively spliced isoforms."  
 RL Genomics 50:129-146(1998).  
 DR EMBL: AF062318; AAC39803.1; -  
 DR InterPro: IPR000225; -  
 DR FLYBase: FBgn000308; -  
 DR Pfam: PF00514; Armadillo\_seg; 5.

DR PRODM: PD000600; -: 1.  
DR PROSITE: PSS0176; ARM\_REPEAT; 3.  
DR SMART: SM00185; ARM; 1  
SQ SEQUENCE 616 AA; 68851 MW; E1014C183DAFB90D CRC64;

Query Match 0.9%; Score 11; DB 4; Length 616;  
Best Local Similarity 100.0%; Pred. No. 0.06;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 568 SNAAYLQHLG 578  
|||||  
DB 61 SNAAYLQHLG 71

RESULT 15

Q9DE61 PRELIMINARY; PRT; 742 AA.  
AC Q9DE61;  
DT 01-MAR-2001 (TRENBLREL. 16, Created)  
DT 01-MAR-2001 (TRENBLREL. 16, Last sequence update)  
DT 01-MAR-2001 (TRENBLREL. 16, Last annotation update)  
DE P120 (FRAGMENT)  
OS Xenopus laevis (African clawed frog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;  
OC Xenopodinae; Xenopus.  
OX NCBI\_TaxID=8355;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Ciesiolka M., Vanlandschoot A., Staes K., van Roy F.;  
RT "Armadillo-related proteins in Xenopus laevis."  
RL Submitted (May-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF150746; AAC45945.1; -  
FT NON\_TER 1  
SQ SEQUENCE 742 AA; 82927 MW; E51D3A3A5CB95FE7 CRC64;

Query Match 0.9%; Score 11; DB 13; Length 742;  
Best Local Similarity 100.0%; Pred. No. 0.071;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 568 SNAAYLQHLG 578  
|||||  
DB 196 SNAAYLQHLG 206

Search completed: July 19, 2001, 16:11:29  
Job time: 218 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 19, 2001, 16:04:00 ; Search time 36.18 Seconds  
(without alignments)  
4479.648 Million cell updates/sec

Title: US-09-501-171A-4

Perfect score: 6377

Sequence: 1 MFARKPGGAFLGAMPVDDQ.....PYSELNETHSPADPSWV 1225

Scoring table: BLOSUM62

Searched: 425026 seqs, 132305027 residues

Total number of hits satisfying chosen parameters: 425026

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: SP-archaea:\*  
2: SP-bacteria:\*  
3: SP-fungi:\*  
4: SP-human:\*  
5: SP-invertebrate:\*  
6: SP-mammal:\*  
7: SP-mmc:\*  
8: SP-organella:\*  
9: SP-phage:\*  
10: SP-plant:\*  
11: SP-rodent:\*  
12: SP-unclassified:\*  
13: SP-vertebrate:\*  
14: SP-virus:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6377	100.0	1225	4	Q9UB33 homo sapien
2	6344	99.5	1225	4	Q00379 mus musculu
3	6129	96.1	1247	11	Q35927 mus musculu
4	5435	85.2	1040	4	Q9UM66 mus musculu
5	4598	72.1	876	4	Q43840 mus musculu
6	2818.5	44.2	1211	4	Q99569 mus musculu
7	1646.5	25.8	321	4	Q43206 mus musculu
8	1617	25.4	307	4	Q15390 mus musculu
9	1474.5	23.1	907	13	Q9DE92 xenopus lae
10	1459	22.9	742	13	Q9DE61 xenopus lae
11	1372	21.5	264	11	Q35116 rattus norv
12	1371	21.5	939	4	Q60714 mus musculu
13	1371	21.5	939	4	Q60716 mus musculu
14	1371	21.5	939	4	Q60715 mus musculu
15	1370	21.5	939	4	Q60713 mus musculu
16	1362	21.4	941	4	Q9UP71 mus musculu
17	1288	20.2	838	4	Q9UP71 mus musculu
18	1288	20.2	885	4	Q9UP72 mus musculu

20	1256	19.7	781	5	Q9NHP1	Q9NHP1 drosophila
21	1245.5	19.5	616	4	Q9UP73	Q9UP73 homo sapien
22	1171.5	18.4	779	5	Q9M579	Q9M579 drosophila
23	954.5	15.0	1181	5	Q9U308	Q9U308 caenorhabd
24	833	13.1	837	4	Q99960	Q99960 homo sapien
25	801	12.6	881	4	Q99959	Q99959 homo sapien
26	770.5	12.1	797	4	Q9Y446	Q9Y446 homo sapien
27	766	12.0	797	11	Q9OY23	Q9OY23 mus musculu
28	719.5	11.3	725	4	Q15152	Q15152 homo sapien
29	713	11.2	725	4	Q13835	Q13835 homo sapien
30	709.5	11.1	728	11	P97350	P97350 mus musculu
31	689	10.8	727	6	Q28161	Q28161 bos taurus
32	595	9.3	114	4	Q9UPW3	Q9UPW3 bos taurus
33	504	7.9	295	6	Q28875	Q28875 bos taurus
34	286	4.5	91	4	Q95645	Q95645 homo sapien
35	267.5	4.2	1566	4	Q9P286	Q9P286 homo sapien
36	266.5	4.2	995	4	Q9Y2W4	Q9Y2W4 homo sapien
37	266.5	4.2	1012	4	Q75359	Q75359 homo sapien
38	257	4.0	1190	4	Q99621	Q99621 homo sapien
39	256.5	4.0	1268	4	Q75046	Q75046 homo sapien
40	255.5	4.0	1012	4	Q43393	Q43393 homo sapien
41	255.5	4.0	1175	11	P70200	P70200 mus musculu
42	255.5	4.0	1182	4	Q99495	Q99495 homo sapien
43	245.5	3.8	1175	11	Q35126	Q35126 mus musculu
44	232.5	3.6	1006	11	Q62901	Q62901 rattus norv
45	224.5	3.5	3938	11	Q88778	Q88778 rattus norv

## ALIGNMENTS

RESULT	1	
ID	Q9UB33	PRELIMINARY; PRT; 1225 AA.
AC	Q9UB33	
DT	01-MAY-2000 (TREMUR1.13, Created)	
DT	01-MAY-2000 (TREMUR1.13, Last sequence update)	
DT	01-MAR-2001 (TREMUR1.16, Last annotation update)	
DE	NEURAL PLAKOPHILIN-RELATED ARM-REPEAT PROTEIN (NRPAP).	
OS	Homo sapiens (Human).	
CC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
CC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
OX	NCBI_TaxID=9606;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=BRNIN;	
RX	MEDLINE-99222289; PubMed-10208590;	
RA	Tanahashi H.;	
RT	Isolation of human delta-catenin and its binding specificity with	
RL	Neuroreport 10:563-568(1999).	
DR	EMBL: AB013805; BAA36163.1; -	
DR	InterPro: IPR000225;	
DR	Pfam: PF00514; Armadillo_seg; 6.	
DR	PROSITE: PS50176; ARM_REPEAT; 3.	
DR	SMART: SM00185; ARM; 1.	
SO	SEQUENCE 1225 AA; 132655 MW; 4A24ACC3E22BFE83 CRC64;	

2-5-1999

Query Match	100.0%	Score 6377;	DB 4;	Length 1225;
Best Local Similarity	100.0%	Pred. No. 0;		
Matches 1225;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;
QY	1	MFARKPGGAFLGAMPVDDQSSASEKTSLSPLGNTSNGDSEETTTSTIILASVKEQEL	60	
DB	1	MFARKPGGAFLGAMPVDDQSSASEKTSLSPLGNTSNGDSEETTTSTIILASVKEQEL	60	
QY	61	QFERLRLLEAEROIVASQERCKLGSSTGSSMSAEEQFOWQSDQDQKIDEDLTG	120	
DB	61	QFERLRLLEAEROIVASQERCKLGSSTGSSMSAEEQFOWQSDQDQKIDEDLTG	120	
QY	121	LELVDSICRSIQESGILDPQDYSGERPSLLSQAOLNKPGRGSPYPSVSYNOTIAL	180	
DB	121	LELVDSICRSIQESGILDPQDYSGERPSLLSQAOLNKPGRGSPYPSVSYNOTIAL	180	

```

Db 121 LELVDSICIRLSQESGILDPDYSTGERPSLLSOSALQOLNSKREGSFQYPASVHNSQTLAL 180
QY 181 GETTPSOILPARGTOARATGOSFSOGTTSRAGHLAGEPAPPPPPPPPPPPPPPPPSIGSAFHL 240
Db 181 GETTPSOILPARGTOARATGOSFSOGTTSRAGHLAGEPAPPPPPPPPPPPPPPPPSIGSAFHL 240
QY 241 PDAPPAALAAALYYSSSTLPAPPPGSGPLAPPOGSGPTKLORGSGAEGATYAAPRGSSP 300
Db 241 PDAPPAALAAALYYSSSTLPAPPPGSGPLAPPOGSGPTKLORGSGAEGATYAAPRGSSP 300
QY 301 KOSPSRLAKSYSTSPINIVYSSAGLSPIRYTSPPTVOSTISSPPIHOLSSSTIGTYATLS 360
Db 301 KOSPSRLAKSYSTSPINIVYSSAGLSPIRYTSPPTVOSTISSPPIHOLSSSTIGTYATLS 360
QY 361 PTKRLVHASEOYKSHSOELVATLQORGSALAAGSRASYSQOHGLGPDELALQSPPHI 420
Db 361 PTKRLVHASEOYKSHSOELVATLQORGSALAAGSRASYSQOHGLGPDELALQSPPHI 420
QY 421 DEIVEDRYOKRPMKSLQSOGDPLPRAHTGYRTSTAPSSPGVDVSVPLORTGSQHPON 480
Db 421 DEIVEDRYOKRPMKSLQSOGDPLPRAHTGYRTSTAPSSPGVDVSVPLORTGSQHPON 480
QY 481 AAAATFOAASVYAGPASYADPYRQLOVCPSESPYKSGALPREGTLASPSIDSIQ 540
Db 481 AAAATFOAASVYAGPASYADPYRQLOVCPSESPYKSGALPREGTLASPSIDSIQ 540
QY 541 DPREGMDPELPEYIOMLOHOPPSVOSNAAYLOHLCFGDNKIKAEIRROGQIQLVLD 600
Db 541 DPREGMDPELPEYIOMLOHOPPSVOSNAAYLOHLCFGDNKIKAEIRROGQIQLVLD 600
QY 601 LDHRTVEVHRSACGALRLVYKANDNKIALKNCGGIPALVRLRLRTTDEIRLEVTV 660
Db 601 LDHRTVEVHRSACGALRLVYKANDNKIALKNCGGIPALVRLRLRTTDEIRLEVTV 660
QY 661 LMNLSGCCALMPITQDLAVLTNAVITPHSGWENSPLODRKTIQLHSSQVLRNATGCLR 720
Db 661 LMNLSGCCALMPITQDLAVLTNAVITPHSGWENSPLODRKTIQLHSSQVLRNATGCLR 720
QY 721 NVSSAGEBARRMRECDLTPALVYIOSALGSSPIDSKTYENCVCILRLNSTYLAETS 780
Db 721 NVSSAGEBARRMRECDLTPALVYIOSALGSSPIDSKTYENCVCILRLNSTYLAETS 780
QY 781 OGOHMGTDLDGLCGEANGDAESSGCGKRRKKSQDQMDVGPLDCAEPPKGIOM 840
Db 781 OGOHMGTDLDGLCGEANGDAESSGCGKRRKKSQDQMDVGPLDCAEPPKGIOM 840
QY 841 LMHPSIVRPIYTLTLLSECSNPDTLEGAAGALQNLAAAGSWKWSVYIRAAVRKKGKGLPIVEL 900
Db 841 LMHPSIVRPIYTLTLLSECSNPDTLEGAAGALQNLAAAGSWKWSVYIRAAVRKKGKGLPIVEL 900
QY 901 LRINDRVYCAVATLRNALMDVNRKELIGRYAMRDVLRPLPGNNSNNTASKAMSDTV 960
Db 901 LRINDRVYCAVATLRNALMDVNRKELIGRYAMRDVLRPLPGNNSNNTASKAMSDTV 960
QY 961 TAVCTLEHVITTKNNENAKALRDAGIEKLVGSKSGDKHSPPVYKAAOVLNMSMOYR 1020
Db 961 TAVCTLEHVITTKNNENAKALRDAGIEKLVGSKSGDKHSPPVYKAAOVLNMSMOYR 1020
QY 1021 DLRSLYKKGDSOYHFAVSSSTIERDORPYSSSRTPSISVRYSPNNRSASAPSPREM 1080
Db 1021 DLRSLYKKGDSOYHFAVSSSTIERDORPYSSSRTPSISVRYSPNNRSASAPSPREM 1080
QY 1081 ISLKERKTDYECTGSNATYHGAKEHTSRKDAANTQNTGISTLYRNSYGAAEDIKHNOY 1140
Db 1081 ISLKERKTDYECTGSNATYHGAKEHTSRKDAANTQNTGISTLYRNSYGAAEDIKHNOY 1140
QY 1141 SAQVPYQEPSRKDYETTYPFONSTRANDSEFFEDQVHHRPPASEYTMHLGKSGNTYVD 1200
Db 1141 SAQVPYQEPSRKDYETTYPFONSTRANDSEFFEDQVHHRPPASEYTMHLGKSGNTYVD 1200
QY 1201 YSAARPYSELNETHSHYPASPSDW 1225
Db 1201 YSAARPYSELNETHSHYPASPSDW 1225

```

```

RESULT 2
ID 000379 PRELIMINARY; PRT: 1225 AA.
AC 000379; Q13589;
DT 01-JUL-1997 (TREMblrel. 04, Created)
DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
DT 01-MAR-2001 (TREMblrel. 16, Last annotation update)
DE DELTA-CATENIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN 11
RP SEQUENCE FROM N.A.
KA MEDLINE=97366296; PubMed=9223106;
RA Zhou J., Liyanage U., Medina M., Ho C., Simmons A.D., Kosik K.S.;
RT "Presentation of interaction in the brain with a novel member of the
RT Armadillo family."
RL Neuroreport 8:2085-2090(1997).
RN 12
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RA Lu Q., Paredes M., Medina M., Zhou J., Cavallo R., Peifer M.,
RA Orecchio L., Kosik K.S.;
RT "d-catenin, an adhesive junction associated protein which promotes
RT motile behavior."
RL Submitted (Oct-1998) to the EMBL/Genbank/DBJ databases.
RN 13
RP SEQUENCE OF 865-1225 FROM N.A.
RX MEDLINE=97202103; PubMed=9049630;
RA Simmons A.D., Overhauser J., Lovett M.;
RT "Isolation of cDNAs from the Cri-du-chat critical region by direct
RT screening of a chromosome 5 specific cDNA library."
RL Genome Res. 7:118-127(1997).
DR EMBL; 090197; AAC00409.1;
DR InterPro; IPR000225;
DR Pfam; PF00514; Armadillo_seg; 6.
DR PROSITE; PSS0176; ARM_REPEAT; 3.
DR SMART; SM00185; ARM; 1.
SQ SEQUENCE 1225 AA; 132664 MW; 8B676CDD5AFAE65 CRC64;

Query Match 99.5%; Score 6344; DB 4; Length 1225;
Best local Similarity 99.6%; Pred. No. 0;
Matches 1220; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```

```

QY 361 PTKRLVHASEQYKSHQSOELVATATLQRPGLSLAAGRASYSOHHGLGPELRALQSPENHI 420
    |||
Db 361 PTKRLVHASEQYKSHQSOELVATATLQRPGLSLAAGRASYSOHHGLGPELRALQSPENHI 420
QY 421 DPTIEDRYVOKPQPMKSLQSOQDPLPRAHTGYRTSTAPSSGVDVSVLQRTGSOHGPON 480
    |||
Db 421 DPTIEDRYVOKPQPMKSLQSOQDPLPRAHTGYRTSTAPSSGVDVSVLQRTGSOHGPON 480
QY 481 AAAATFORASVYAGPASNADYRQLOQCPSEVSPSKSGPALPEGTARSPTSIDSIOK 540
    |||
Db 481 AAAATFORASVYAGPASNADYRQLOQCPSEVSPSKSGPALPEGTARSPTSIDSIOK 540
QY 541 DPREFGMRDPELPEVYIOMLQHOFPVSQNAAYLQHLCPGDKIKAEIRROGGIQLVLDL 600
    |||
Db 541 DPREFGMRDPELPEVYIOMLQHOFPVSQNAAYLQHLCPGDKIKAEIRROGGIQLVLDL 600
QY 601 LDHRTVEVHRSCAGLRNLVYGKANDNKIALKNCGGIPALVRLRKTDTLEIRELVYGV 660
    |||
Db 601 LDHRTVEVHRSCAGLRNLVYGKANDNKIALKNCGGIPALVRLRKTDTLEIRELVYGV 660
QY 661 LWNLSGCDALKMPITIDALAVLTNAVITPHSGWENSPLODDKRIOLHSSQVLRNATGCLR 720
    |||
Db 661 LWNLSGCDALKMPITIDALAVLTNAVITPHSGWENSPLODDKRIOLHSSQVLRNATGCLR 720
QY 721 NVSSAGEARRMRRCDCGLTDALVYIOLGALSSEIDSKTVENCYILRNLSYRLAETS 780
    |||
Db 721 NVSSAGEARRMRRCDCGLTDALVYIOLGALSSEIDSKTVENCYILRNLSYRLAETS 780
QY 781 QOGHMTDELIDGLGCEANGKDAESSGCKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 840
    |||
Db 781 QOGHMTDELIDGLGCEANGKDAESSGCKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 840
QY 841 LHMPSIVKPLTLLSCSNPDTEGAAGALONLAAGSMKMSYIRAAVKEKGLPLTVEL 900
    |||
Db 841 LHMPSIVKPLTLLSCSNPDTEGAAGALONLAAGSMKMSYIRAAVKEKGLPLTVEL 900
QY 901 LRIDNRVCAVATALRNALDVRNKLIGKAYMRDLVHRLPGGNSNNTASKAMSDDTY 960
    |||
Db 901 LRIDNRVCAVATALRNALDVRNKLIGKAYMRDLVHRLPGGNSNNTASKAMSDDTY 960
QY 961 TAVCCTLHEVITTKMENAKALRDAGIEKLVGSKSGKHSKPKVYKAAQVLYNSMQR 1020
    |||
Db 961 TAVCCTLHEVITTKMENAKALRDAGIEKLVGSKSGKHSKPKVYKAAQVLYNSMQR 1020
QY 1021 DLMSLKQKQMSQHYHVAASSTIERDORPYSSTPISPVSPNNNSASAPASPRFM 1080
    |||
Db 1021 DLMSLKQKQMSQHYHVAASSTIERDORPYSSTPISPVSPNNNSASAPASPRFM 1080
QY 1081 ISLKERKTDYECTGSNATYHGAKEHTSRKDAVTAONTGISTLYRNSYGAPADIKHNOY 1140
    |||
Db 1081 ISLKERKTDYECTGSNATYHGAKEHTSRKDAVTAONTGISTLYRNSYGAPADIKHNOY 1140
QY 1141 SAQVPOEPRKRYETIYOPQONSTRNYDESEFFDOYHHRPASEYTMHLGKSTGYVDF 1200
    |||
Db 1141 SAQVPOEPRKRYETIYOPQONSTRNYDESEFFDOYHHRPASEYTMHLGKSTGYVDF 1200
QY 1201 YSARPYSELNTEYSHYPASPSMV 1225
    |||
Db 1201 YSARPYSELNTEYSHYPASPSMV 1225

```

## RESULT 3

035927 PRELIMINARY: PRT: 1247 AA.

```

AC 035927;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, last annotation update)
DE NEURBL PLAKOPHILIN-RELATED ARM-REPEAT PROTEIN.
GN CTRNND2 OR NRAP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```

```

OX NCBI TaxID=10090;
RN 11
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN.
RX MEDLINE=98002299; PubMed=9342840;
RA Patenholz R., Franke W.W.;
RT "Identification and localization of a neurally expressed member of the
RL plakoglobin/armadillo multigene family.";
DR EMBL: U90331; AAB82409.1;
DR MGP: MGI:1195966; Ctrnd2.
DR InterPro: IPR000225;
DR Pfam: PF00514; Armadillo_seg; 6.
DR PROSITE: PS0176; ARM_REPEAT; 3.
DR SMART: SM00185; ARM; 1.
SQ SEQUENCE 1247 AA; 134998 MW; DAA7A6B6A27D2919 CRC64;

Query Match 96.1%; Score 6129; DB 11; Length 1247;
Best Local Similarity 94.8%; Pred. No. 0;
Matches 1185; Conservative 8; Mismatches 29; Indels 28; Gaps 3;

QY 1 MPARKPEAAALGAMPVPDOPSSASEKTSLSPLGNTSNGDSTETTSAILASVKEQL 60
    |||
Db 1 MPARKOSGAAPGAMPVPDOPPSASEKNSLSPLGNTSNGDSTETTSAILASVKEQL 60
QY 61 QPERLTRELAEPOIVASQLECKIGSEFTGSMSSSAEBOFQWQODGQKDIJEDLTG 120
    |||
Db 61 QPERLTRELAEPOIVASQLECKIGSEFTGSMSSSAEBOFQWQODGQKDIJEDLTG 120
QY 121 LEVYDSCIRLSQESGILDPDQYSTEGERPSLLSQSALQNSKEGSEFOYPASVSNQTLAL 180
    |||
Db 121 LEVYDSCIRLSQESGILDPDQYSTEGERPSLLSQSALQNSKEGSEFOYPASVSNQTLAL 180
QY 181 GETTSPQAPAGTOARATGQSFSGTTSBRAGHLAGPEAPPPPPPPPPPPPPPPPPPPPP 240
    |||
Db 181 GDTASQPLPASTQARAQGSFSGTTSBRAGHLAGSEPA-PPPPPPPPPPPPPPPPPPPP 239
QY 241 PDAPPAALAAALYSSSTLPAPPRGSPPLAPOGSPTKLQRGSAPEGATYAAPRGSSP 300
    |||
Db 240 PDAPP--AAALYSSSTLPAPPRGSPPLTTOGGSPTKLQRGSAPEGATYAAPRGSSP 297
QY 301 KQSPRLAKSYSTSSPINIVVSSAGLSPIRVTSPPTVOOSTSSSPHOLSTIGTYATLS 360
    |||
Db 298 KQSPRLAKSYSTSSPINIVVSSAGLSPIRVTSPPTVOOSTSSSPHOLSTIGTYATLS 357
QY 361 PTKRLVHASEQYKSHQSOELVATATLQRPGLSLAAGRASYSOHHGLGPELRALQSPENHI 420
    |||
Db 358 PTKRLVHASEQYKSHQSOELVATATLQRPGLSLAAGRASYSOHHGLGPELRALQSPENHI 417
QY 421 DPTIEDRYVOKPQPMKSLQSOQDPLPRAHTGYRTSTAPSSGVDVSVLQRTGSOHGPON 480
    |||
Db 418 DPTIEDRYVOKPQPMKSLQSOQDPLPRAHTGYRTSTAPSSGVDVSVLQRTGSOHGPON 477
QY 481 AAAATFORASVYAGPASNADYRQLOQCPSEVSPSKSGPALPEGTARSPTSIDSIOK 540
    |||
Db 478 AAAATFORASVYAGPASNADYRQLOQCPSEVSPSKSGPALPEGTARSPTSIDSIOK 537
QY 541 DPREFGMRDPELPEVYIOMLQHOFPVSQNAAYLQHLCPGDKIKAEIRROGGIQLVLDL 600
    |||
Db 538 DPREFGMRDPELPEVYIOMLQHOFPVSQNAAYLQHLCPGDKIKAEIRROGGIQLVLDL 597
QY 601 LDHRTVEVHRSCAGLRNLVYGKANDNKIALKNCGGIPALVRLRKTDTLEIRELVYGV 660
    |||
Db 598 LDHRTVEVHRSCAGLRNLVYGKANDNKIALKNCGGIPALVRLRKTDTLEIRELVYGV 657
QY 661 LWNLSGCDALKMPITIDALAVLTNAVITPHSGWENSPLODDKRIOLHSSQVLRNATGCLR 720
    |||
Db 658 LWNLSGCDALKMPITIDALAVLTNAVITPHSGWENSPLODDKRIOLHSSQVLRNATGCLR 717
QY 721 NVSSAGEARRMRRCDCGLTDALVYIOLGALSSEIDSKTVENCYILRNLSYRLAETS 780
    |||
Db 718 NVSSAGEARRMRRCDCGLTDALVYIOLGALSSEIDSKTVENCYILRNLSYRLAETS 777

```

```

QY 781 OQOHMGTELDGLGGEANGKDAESSGCGKKKKKKSSODOMDGVGRLPDCAEPPKGIOM 840
DB 778 OQOHMGTELDGLGGEANGKDAESSGCGKKKKKKSSODOMDGVGRLPDCAEPPKGIOM 837
QY 841 LMHPSIVKRYLLTLLSECSNPDTLEGAAGALONLAAGSWK----- 879
DB 838 LMHPSIVKRYLLTLLSECSNPDTLEGAAGALONLAAGSWKMAEDVAGMAVALRLSPGAP 897
QY 880 ----WSYIRAAVREKKEGPIVELLRIDNDRVCAVATLRNALDVNRKELLGKIYMR 935
DB 898 CLPQMSYVIRAAVREKKEGPIVELLRIDNDRVCAVATLRNALDVNRKELLGKIYMR 957
QY 936 DEVHRLPGGNNNTASRAMSDTYTAVCCTLHEVITKNMNAKALRPAAGIEKLYGISK 995
DB 958 DLVHRLPGGNNNTASRAMSDTYTAVCCTLHEVITKNMNAKALRPAAGIEKLYGISK 1017
QY 996 SKGDHSPKRYVKAASQVLSNMWQYRDLRSILYKKDQMSQYHEVASSSTIERDORPYSRR 1055
DB 1018 SKGDHSPKRYVKAASQVLSNMWQYRDLRSILYKKDQMSQYHEVASSSTIERDORPYSRR 1077
QY 1096 TPISIPVSPVSPNNRSASAPSPREMI SLKERTDYECTGSNATYHGAKGERTSKRDAMTA 1115
DB 1078 TPISIPVSPVSPNNRSASAPSPREMI SLKERTDYESAGNNATYHGTEHTSKRDTMTA 1137
QY 1116 QNTGISTLYRNSYGAPADIDIKNOVSAOPVPOEPSRKDYETIYOPFONSTRNYDESFEDQ 1175
DB 1138 QNTGISTLYRNSYGAPADIDIKNOVSTOVPPOEPSRKDYETIYOPFONSTRNYDESFEDQ 1197
QY 1176 VHHRRPASEYTMHGLKSTGNYVDFYSARPYSELNETHYSPASPSDW 1225
DB 1198 VHHRRPASEYTMHGLKSTGNYVDFYSARPYSELNETHYSPASPSDW 1247

RESULT 4
QY 090M66 PRELIMINARY; PRT: 1040 AA.
AC 090M66;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)
DE C724 (FRAGMENT).
GN C724.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
CX NCBI_TaxID=9606;
RN 111
RP SEQUENCE FROM N.A.
RA Levesque G., Yu G., Fraser P.E., StGeorge-Hyslop P.;
RT "Presentin 1 interacts with a novel protein that contains armadillo
RL submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: U81004; AAD00453.1; -
DR InterPro: IPR000225; -
DR Pfam: PF00514; Armadillo_seg. 6.
DR SMART: PS50176; ARM_REPEAT; 3.
DR SMART: SM00185; ARM; 1.
FT NON_TER 1
FT SEQUENCE 1040 AA; 112802 MW; E12C8694EAC51D42 CXC64;

Query Match 85.2%; Score 5435; DB 4; Length 1040;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1037; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```

```

QY 306 RLAKSYSTSPINIVYSSAGLSPIRVTSPPTVOSTISSPFIHQLSTICTYATLSPTRL 365
DB 121 RLAKSYSTSPINIVYSSAGLSPIRVTSPPTVOSTISSPFIHQLSTICTYATLSPTRL 180
QY 366 VHASEOYKSHQSLQELYTATLQRPGLAAGSRASYSOHLGPELALQSPENHIDPYE 425
DB 181 VHASEOYKSHQSLQELYTATLQRPGLAAGSRASYSOHLGPELALQSPENHIDPYE 240
QY 426 DRYVQRPMSLSLSQSGDPLPRAHTGTYRTSTAPSSPGVDSVPLQRTGSOHGPONAAAT 485
DB 241 DRYVQRPMSLSLSQSGDPLPRAHTGTYRTSTAPSSPGVDSVPLQRTGSOHGPONAAAT 300
QY 486 FQASVTAAGASVYADPYRQLODYCEVSEPSYKSGPALPEGLTARSPSIDIQDPRF 545
DB 301 FQASVTAAGASVYADPYRQLODYCEVSEPSYKSGPALPEGLTARSPSIDIQDPRF 360
QY 546 GMRDPELPEYIOMLQHPFVSQSNAAAYLOHLCFGDNKTKAETIRROGGTQLVLDLDRM 605
DB 361 GMRDPELPEYIOMLQHPFVSQSNAAAYLOHLCFGDNKTKAETIRROGGTQLVLDLDRM 420
QY 606 TEVHRSACGALRLVYKANDDNKIALKNCGGIPALVRLRKTDTLEIRELYGYLMLNS 665
DB 421 TEVHRSACGALRLVYKANDDNKIALKNCGGIPALVRLRKTDTLEIRELYGYLMLNS 480
QY 666 SCQALMPTIIOALAVITNAVITPHSGWENSPLODDRKTQLHSSQVLRNATGCLRNSSA 725
DB 481 SCQALMPTIIOALAVITNAVITPHSGWENSPLODDRKTQLHSSQVLRNATGCLRNSSA 540
QY 726 GEBARRMECDLTPALVYIQLSALGSEISDKYENCVCILRLNLSYLAETISQOGHM 785
DB 541 GEBARRMECDLTPALVYIQLSALGSEISDKYENCVCILRLNLSYLAETISQOGHM 600
QY 786 GTDELDGLGGEANGKDAESSGCGKKKKKKSSODOMDGVGRLPDCAEPPKGIOMLHPS 845
DB 601 GTDELDGLGGEANGKDAESSGCGKKKKKKSSODOMDGVGRLPDCAEPPKGIOMLHPS 660
QY 846 IYKPYTLTLLSECSNPDTLEGAAGALONLAAGSWKMSVYIRAAVREKKEGPIVELLRIDN 905
DB 661 IYKPYTLTLLSECSNPDTLEGAAGALONLAAGSWKMSVYIRAAVREKKEGPIVELLRIDN 720
QY 906 DRYVCAVATLRNALDVNRKELLGKIYMRDLVHRLPGGNNNTASRAMSDTYTAVCC 965
DB 721 DRYVCAVATLRNALDVNRKELLGKIYMRDLVHRLPGGNNNTASRAMSDTYTAVCC 780
QY 966 TLHEVITKNMNAKALRDAAGIEKLYGISKSGDHSKRYVKAASQVLSNMWQYRDLRS 1025
DB 781 TLHEVITKNMNAKALRDAAGIEKLYGISKSGDHSKRYVKAASQVLSNMWQYRDLRS 840
QY 1026 YKKGMSQYHEVASSSTIERDORPYSRRTPISPVRSPPNNRSASAPSPREMI SLKE 1085
DB 841 YKKGMSQYHEVASSSTIERDORPYSRRTPISPVRSPPNNRSASAPSPREMI SLKE 900
QY 1086 RKTDECTGSNATYHGAKGERTSKRDAMTAQNTGISTLYRNSYGAPADIDIKNOVSAOPV 1145
DB 901 RKTDECTGSNATYHGAKGERTSKRDAMTAQNTGISTLYRNSYGAPADIDIKNOVSAOPV 960
QY 1146 POEPSRKDYETIYOPFONSTRNYDESFEDQVHHRRPASEYTMHGLKSTGNYVDFYSAR 1205
DB 961 POEPSRKDYETIYOPFONSTRNYDESFEDQVHHRRPASEYTMHGLKSTGNYVDFYSAR 1020
QY 1206 PYSELNETHYSPASPSDW 1225
DB 1021 PYSELNETHYSPASPSDW 1040

RESULT 5
QY 043840 PRELIMINARY; PRT: 876 AA.
AC 043840;
DT 01-JUN-1998 (Tremblrel. 06, Created)
DT 01-JUN-1998 (Tremblrel. 06, Last sequence update)
DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)

```

DE NEURAL PLAKOPHILIN RELATED ARM-REPEAT PROTEIN (FRAGMENT).  
 GN NRPAP.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=BRNIN;  
 RX MEDLINE=98002299; PubMed=9342840;  
 RA Patenholz R., Franke W.W.;  
 RT "Identification and localization of a neurally expressed member of the  
 RT plakoglobin/armadillo multigenic family";  
 RL Differentiation 61:293-304(1997).  
 DR EMBL: U52351; AAB97957.1; -  
 DR InterPro: IPR000225; -  
 DR Pfam: PF00514; Armadillo\_seg; 6.  
 DR PROSITE: PS0176; ARM\_REPEAT; 3.  
 DR SMART: SM00185; ARM; 1.  
 FT NON\_TER  
 FT 1  
 SO SEQUENCE 876 AA; 96442 MW; 8D342B71C9E92A CRC64;

Query Match 72.1%; Score 4598; DB 4; Length 876;  
 Best Local Similarity 99.9%; Pred. No. 5.3e-292;  
 Matches 875; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 350 SSTIGYATLSTPKRLVHASEOYKSHOELVATLQPGSLAASRAYSOSHGLPE 409  
 DB 1 SSTIGYATLSTPKRLVHASEOYKSHOELVATLQPGSLAASRAYSOSHGLPE 60  
 OY 410 LRALOSPEHHIDPIEDRYOKRPMKSLQSOGDPLPPAHGTGYRTSTAPSPGVDSVPL 469  
 DB 61 LRALOSPEHHIDPIEDRYOKRPMKSLQSOGDPLPPAHGTGYRTSTAPSPGVDSVPL 120  
 OY 470 QRTSGHGHGQNAATAATFQAASTAGPASYADPYQOLQYCEVESPYKSGALPPEGL 529  
 DB 121 QRTSGHGHGQNAATAATFQAASTAGPASYADPYQOLQYCEVESPYKSGALPPEGL 180  
 OY 530 ARSPSTDSIOKPRPFGMDPELPEYIOMLQHPFVSQNAAYLOHLCFQGNKTKAEIR 589  
 DB 181 ARSPSTDSIOKPRPFGMDPELPEYIOMLQHPFVSQNAAYLOHLCFQGNKTKAEIR 240  
 OY 590 RGGIGDLDVLDLHRTVEHRSACGALRNLYGKANDNKIALKNCGGIPALVRLRKT 649  
 DB 241 RGGIGDLDVLDLHRTVEHRSACGALRNLYGKANDNKIALKNCGGIPALVRLRKT 300  
 OY 650 DLEIRLVGVLMNLSGCCALKPITIDALAVLTNAVITPHSGWENSPLODRKIQHSS 709  
 DB 301 DLEIRLVGVLMNLSGCCALKPITIDALAVLTNAVITPHSGWENSPLODRKIQHSS 360  
 OY 710 QVLRNATGCLRNVSSAGEARRRRCDDLTJALLVIOISAGSSSIDSKTYENCYILR 769  
 DB 361 QVLRNATGCLRNVSSAGEARRRRCDDLTJALLVIOISAGSSSIDSKTYENCYILR 420  
 OY 770 NLSYRLAETSGOQHMGTDLDLGLCGEANGKDAESSGGCKKKKKKQSDQMDGVPLP 829  
 DB 421 NLSYRLAETSGOQHMGTDLDLGLCGEANGKDAESSGGCKKKKKKQSDQMDGVPLP 480  
 OY 830 DCAEPKGIOMLHPSIVRYLTLLSECSNPDTLBSAQAQONLAAGSKKSVYIRAAVR 889  
 DB 481 DCAEPKGIOMLHPSIVRYLTLLSECSNPDTLBSAQAQONLAAGSKKSVYIRAAVR 540  
 OY 890 KEKGLLIVELLIRINDRYVCAVATLRNALDVRKELIGYAMDLVHRLPGGNNSNN 949  
 DB 541 KEKGLLIVELLIRINDRYVCAVATLRNALDVRKELIGYAMDLVHRLPGGNNSNN 600  
 OY 950 TASKAMSDDTVTAVCTLHEVITKKNENAKALDAAGIEKLVGISKSGDKSPKVVKA 1009  
 DB 601 TASKAMSDDTVTAVCTLHEVITKKNENAKALDAAGIEKLVGISKSGDKSPKVVKA 660  
 OY 1010 SOVLNSMWQYRDLRLSLYKKDGWQYHFVASSSTIERDRRPYSSSTPSTISYRVSPNNR 1069  
 DB 1010 SOVLNSMWQYRDLRLSLYKKDGWQYHFVASSSTIERDRRPYSSSTPSTISYRVSPNNR 1069

DB 661 SOVLNSMWQYRDLRLSLYKKDGWQYHFVASSSTIERDRRPYSSSTPSTISYRVSPNNR 720  
 OY 1070 SASAPASPREMISLKEKTDYECTGSNATYHGAKGHTSRKDMATQNTGISTLYRNSYG 1129  
 DB 721 SASAPASPREMISLKEKTDYECTGSNATYHGAKGHTSRKDMATQNTGISTLYRNSYG 780  
 OY 1130 APAEDIKHNQVSAQVPVQPSRKDYETQFONSTNRYDSDFEDOVHHRPASERTMIL 1189  
 DB 781 APAEDIKHNQVSAQVPVQPSRKDYETQFONSTNRYDSDFEDOVHHRPASERTMIL 840  
 OY 1190 GLKSTGYVDFYSARPYSELNETHYPASDPDW 1225  
 DB 841 GLKSTGYVDFYSARPYSELNETHYPASDPDW 876

RESULT 6  
 ID 099569 PRELIMINARY; PRT; 1211 AA.  
 AC 099569;  
 DT 01-MAY-1997 (Tremblrel. 03, Created)  
 DT 01-MAY-1997 (Tremblrel. 03, Last sequence update)  
 DT 01-MAR-2001 (Tremblrel. 16, Last annotation update)  
 DE P0071 PROTEIN.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=FRONTAL CORTEX;  
 RX MEDLINE=97092329; PubMed=8937994;  
 RA Hatzfeld M., Nachtsheim C.;  
 RT "Cloning and characterization of a new armadillo family member, p0071,  
 RT associated with the junctional plaque: evidence for a subfamily of  
 RT closely related proteins.";  
 RL J. Cell Sci. 109:2767-2778(1996).  
 DR EMBL: X81889; CAA57478.1; -  
 DR InterPro: IPR000225; -  
 DR Pfam: PF00514; Armadillo\_seg; 7.  
 DR PROSITE: PS0176; ARM\_REPEAT; 3.  
 DR SMART: SM00185; ARM; 1.  
 SO SEQUENCE 1211 AA; 134268 MW; 14C9626914A7024F CRC64;

Query Match 44.2%; Score 2818.5; DB 4; Length 1211;  
 Best Local Similarity 49.2%; Pred. No. 1.3e-175;  
 Matches 618; Conservative 192; Mismatches 337; Indels 109; Gaps 37;

OY 15 MPVPDQSSASEKTSISLPGINTSNGSGSETTT-SAILASVKEQLQFRLTRELAEAR 73  
 DB 1 MPVPDQSSASEKTSISLPGINTSNGSGSETTT-SAILASVKEQLQFRLTRELAEAR 59  
 OY 74 QIVASOLEKRLGSETGSMSSSAEFOFQWOSOD-----GOKDIDELFTTGLLELVDS 127  
 DB 60 QIVASOLEKRLGSETGSMSSSAEFOFQWOSOD-----GOKDIDELFTTGLLELVDS 115  
 OY 128 IRSLOESGIL-DPODYSTERPDL--SQSALQJNSKEGSGFYQYASVHSNQTAL----- 179  
 DB 116 IRTEPEQGTIVSPQTSLSHESEGLSNGSRSTQNSYSDSGYQFAGSFHNSQWNSKADNR 175  
 OY 180 -----LGETTSQLPARKTORAKTGQSFQ-GTTSRAGHLAAGPPAPPP----- 224  
 DB 176 QOHSFISGTTNNHYV-----RNSRAEGQTLVQPSVANRRARVSVPSRAQSPSYISTGVS 221  
 OY 225 PPREPFAISGSAFHLP---DAPPAALAAALYSSSTLPAPPRGSPILAPPOGSPTKLQ 281  
 DB 232 PSRGLRTSLSGSGSPVYDPRPLNPSA---YSTTLPA-ARASPY-SORPASPTAIR 286  
 OY 282 RGSNAPEGATYAADRGSPK-QSPSRLAKSYSTSPINIVSSAGSLPRLVTSPTVQST 340  
 DB 287 RIGSVTSROT-SNPNGPFPQYQTARV-----GSPILTIDAQT-----RVASPSQGO-V 333  
 OY 341 ISSPFIHQLSSTIGYATLSTP-KRLVHASRQYKSHOELVATLQPGSLAAGSASY 399

```

Db 334 GSSSP--KRSGMTAVPQHLGSLQRTVHDMQFGQOQYDIERYVPRPSTL-TGLASSY 330
Qy 400 SSGHGLPELRLALQSPENHIDPIEDRYOKPMPRLSOSQDPLPAITGYRTSTAP 439
Db 391 ASQHSQGLQDRLASVSPDLHTTPTEYGRYYSPYRSPNMGTYE-LOGSOTATLRT---- 445
Qy 460 SSPGVDSV-PLQRTGSGHGPNNAAATFORASVAGASVADRYOLOJOCPSVESYSK 518
Db 446 ---GVSGIGNQRTSSOR-----STLYQNNVYALNTATAYAEYRIRYQRYO-ECYVNR 436
Qy 519 SGPLPPEGTALRSPSIDSTICKPRFEGMDPELPEYIOMLQHPFSVQNAAYIOLHC 578
Db 497 LQHAVPADGDTTRSPSIDSTICKPRFEGMDPELPEYIOMLQHPFSVQNAAYIOLHC 556
Qy 579 FGDKKIAEIRROGIOLVDLIDHMEYHRSAGALRLVYKANKDDKIALKNGGI 638
Db 557 FGDKKIAEIRROGIOLVDLIDHMEYHRSAGALRLVYKANKDDKIALKNGGI 616
Qy 639 PALVRLKRTDLEIRLYTVGLVNLSSCDALKMPITQDALAVLNAVITPHSGWENSPL 698
Db 617 PALVRLKRTDLEIRLYTVGLVNLSSCDALKMPITQDALAVLNAVITPHSGWENSPL 676
Qy 699 QDDKIKIOLHSQVLRNATGCLRNVSAGEARRMRRECDGLTDLALVYIOSALGSEIDS 758
Db 677 DDDKIKIOLHSQVLRNATGCLRNVSAGEARRMRRECDGLTDLALVYIOSALGSEIDS 736
Qy 759 KTVENCYILRLNLSYRLAETSSOGHMGDELGLGCEANGDAESSGCKKKKKKK 817
Db 737 KTVENCYILRLNLSYRLAETSSOGHMGDELGLGCEANGDAESSGCKKKKKKK 795
Qy 818 --SODQMDGVPLDCAEPKGIOMLHPSIVKPYTLTSECSNPDTLEGAGALONLAA 875
Db 796 TPQEDQMDGVPLDCAEPKGIOMLHPSIVKPYTLTSECSNPDTLEGAGALONLAA 855
Qy 876 GSKWASYIIRAAVAKKGLPIVELLRIDNDRYCANATLRMALDYNKELIGYAMR 935
Db 856 SNNFAAYIRGGRKRGPLVELLRMDNDRYVSGATLRMALDYNKELIGYAMR 915
Qy 936 DLVHRLPGNNSNNTASKAMSDTYTAVCTLEHVTIKNNENAKALRDAGIEKLGISK 995
Db 916 DLVHRLPGNNSNNTASKAMSDTYTAVCTLEHVTIKNNENAKALRDAGIEKLGISK 970
Qy 996 SKGDHSPKVVKAASOVLNSMQRDLRLSLYKKGDSQYHFVASSSTIERDORPYSSR 1055
Db 971 GGRGSSSLKAAOVLNLTMQYRLSLYKKGDSQYHFVASSSTIERDORPYSSR 1026
Qy 1056 TPSTSPV--RVSPNNRGSASAPASPREMISLKERDTEGCSNATYHGAKEHTSKDAM 1113
Db 1027 TPSTSPV--RVSPNNRGSASAPASPREMISLKERDTEGCSNATYHGAKEHTSKDAM 1083
Qy 1114 TAONTGISTLYRNSYGAPAD---IKHNOVSAPVPOEPRSKDYETPOFONSRYNDE 1169
Db 1084 YPSSSKSPSYIISYSSPAREQNRRLQHOOLYSS--QODSNRKNFAYRLYLQSPHSYD 1141
Qy 1170 SFPEDOVHHRPASEYTMHGLKSTGNVYDFYGAARYSLANTETSHYSPSPSW 1225
Db 1142 PYFDNRVH--PASTDYSTOYGLKSTGNVYDFYSTKRP---SYRABQYGSPPSW 1192

```

```

RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE=96207227; PubMed=8619474;
RA Andersson B., Wentland M.A., Ricatrente J.Y., Liu W., Gibbs R.A.;
RT "A 'double adaptor' method for improved shotgun library
construction.";
RL Anal. Biochem. 236:107-113(1996).
RN [2].
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE=97264341; PubMed=9110174;
RA Yu W., Andersson B., Worley K.C., Muzny D.M., Ding Y., Liu W.,
RA Ricatrente J.Y., Wentland M.A., Lennon G., Gibbs R.A.;
RT "Large-scale concatenation cDNA sequencing.";
RL Genome Res. 7:353-358(1997).
DR EMBL; AF035302; AAB88185.1;
KW Cytoskeleton; Structural protein; Cell adhesion.
SQ SEQUENCE 321 AA; 36108 MW; 1DDF4811891DA953 CRC64;

```

```

Query Match 25.8%; Score 1646.5; DB 4; Length 321;
Best Local Similarity 98.7%; Pred. No. 7.7e-100;
Matches 315; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

Qy 908 VVCA-VATALRMALDYNKELIGKYARDLVHRLPGNNSNNTASKAMSDTYTAVCCT 966
Db 3 VVCAVSTALRMALDYNKELIGKYARDLVHRLPGNNSNNTASKAMSDTYTAVCCT 62

Qy 967 LHEVITKNNENAKALRDAGIEKLVGISKSKDKHSPKVVKAASOVLNSMQRDLRLSLY 1026
Db 63 LHEVITKNNENAKALRDAGIEKLVGISKSKDKHSPKVVKAASOVLNSMQRDLRLSLY 122

Qy 1027 KKDGMQSYHFAVSSSTIERDORPYSSSRTPSISVPRVSPNNRGSASAPASPREMISLKER 1086
Db 123 KKDGMQSYHFAVSSSTIERDORPYSSSRTPSISVPRVSPNNRGSASAPASPREMISLKER 182

Qy 1087 KTDYECTGSNATYHGAKEHTSRKDMATONTGISTLYRNSYGAPADIEKHNOVSAPV 1146
Db 183 KTDYECTGSNATYHGAKEHTSRKDMATONTGISTLYRNSYGAPADIEKHNOVSAPV 242

Qy 1147 QPSPKQDQETQOPFONSSTRNDESFEEQVYHRRPASYTMHGLKSTGNVDFYSARP 1206
Db 243 QPSPKQDQETQOPFONSSTRNDESFEEQVYHRRPASYTMHGLKSTGNVDFYSARP 302

Qy 1207 YSELNYETSHYPASPSW 1225
Db 303 YSELNYETSHYPASPSW 321

```

```

RESULT 8
O15390
ID O15390 PRELIMINARY; PRT; 307 AA.
AC O15390;
DT 01-JAN-1998 (TREMBlrel. 05, Created)
DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
DT 01-AUG-1998 (TREMBlrel. 07, Last annotation update)
DE G124.
GN G124.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1].
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX Fraser P.E., Levesque G., Rogaeva E.A., Yu G., St George-Hyslop P.H.;
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U72653; AAB6599.1;
SQ SEQUENCE 307 AA; 34417 MW; 35640DA46113B65F CRC64;

```

```

Query Match 25.4%; Score 1617; DB 4; Length 307;
Best Local Similarity 99.7%; Pred. No. 6.1e-98;

```



Query Match	21.5%;	Score 1372;	DB 11;	Length 264;
Best Local Similarity	100.0%;	Pred. No. 5e-82;		
Matches 264;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0
QY 563	FFSVSNAAAYLQHLCHGDNKIKAEIRROGGIQLAVDLIDHRMTEVHNSACGALRNLYVG	622		
DB 1	FFSVSNAAAYLQHLCHGDNKIKAEIRROGGIQLAVDLIDHRMTEVHNSACGALRNLYVG	60		
QY 632	KANDNKTALKNCGGIPALVRLRKTDTLEIRELYGVYMNSSCDALKMPTIIDALAVL	682		
DB 61	KANDNKTALKNCGGIPALVRLRKTDTLEIRELYGVYMNSSCDALKMPTIIDALAVL	120		
QY 683	TNAVLIPIHSGWENSPLODDRRKIQLHSSQVLRNATGCLRNVSSAGEARRRRREDDGLTDA	742		
DB 121	TNAVLIPIHSGWENSPLODDRRKIQLHSSQVLRNATGCLRNVSSAGEARRRRREDDGLTDA	180		
QY 743	LIYIYQSLGSSSEIDSKTVENCYCIIIRNLSTYRLAETSOGCHMTDEIDGLLGEANGKD	802		
DB 181	LIYIYQSLGSSSEIDSKTVENCYCIIIRNLSTYRLAETSOGCHMTDEIDGLLGEANGKD	240		
QY 803	AESSGCWGKK	826		
DB 241	AESSGCWGKK	264		
RESULT 12				
060714	PRELIMINARY:	PR:	939 AA.	
ID 060714				
AD 060714				
DT 01-AUG-1998 (TrEMBLrel. 07, Created)				
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)				
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)				
DE P120 CATENIN ISOFORM 1AC.				
GN CTNND1.				
OS Homo sapiens (Human).				
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC Mammalia; Eutheria; Plimates; Catarrhini; Homnidae; Homo.				
OX NCBI_Taxid=9606;				
RN [1]				
RA SEQUENCE FROM N.A.				
RP Keirnsblck A., Bonne S., Straes K., Van Hengel J., Nollet F.,				
RA Reynolds A., Van Roy F.;				
RL Genomics 0:0-0(1998).				
DR EMBL; AF062321; AAC39806.1; -.				
DR InterPro; IPR000225; -.				
DR Pfam; PF00514; Armadillo_seg. 5.				
DR PROSITE; PSS0176; ARM_REPEAT; 3.				
DR SMART; SM00185; ARM; 1				
SQ SEQUENCE 939 AA; 104976 MW; D436D996CEBBDCT CRC64;				
Query Match	21.5%;	Score 1371;	DB 4;	Length 939;
Best Local Similarity	32.2%;	Pred. No. 3.2e-81;		
Matches 366;	Conservative 165;	Mismatches 322;	Indels 282;	Gaps 31;
QY 41	DGSEFTTSAIIASVKEQLEFERLITRELEAEKQIVASQLECKLG-SETGSMSSMSAE	99		
DB 2	DDSEVESTAIIASVKEQLEAEKFLTRALEERHRSVQALERVSPQDANPIMANGTIL	61		
QY 100	EQRQWQSQDQKRIEDELITGLVDSICRLQESGIIDPDQYSTGRPSILSQSLQLN	159		
DB 62	RRRONGRRVGDALAEKQKPSDLK-----NG---PDHR-----	92		
QY 160	SKEGSRVFPASYHNSNOTALGETTSPQLPARGIQARATQGSFSGTTSRAGHLAGPEPA	219		









Fri Jul 20 10:09:45 2001

us-09-501-171a-4.rpt

---

Page 12

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 19, 2001, 16:00:10 ; Search time 27.68 Seconds

(without alignments)  
2682.962 Million cell updates/sec

Title: US-09-501-171A-4

Perfect score: 6377  
Sequence: 1 MFARKPPGAAPLCAMPVPO.....PSSELYETSHYPASPDSDW 1225

Scoring table:  
BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 412676 seqs, 60623988 residues

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

A.Geneseq 0601.\*  
1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.\*  
2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.\*  
3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.\*  
4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.\*  
5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.\*  
6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.\*  
7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.\*  
8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.\*  
9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.\*  
10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.\*  
11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.\*  
12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.\*  
13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.\*  
14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.\*  
15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.\*  
16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.\*  
17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.\*  
18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.\*  
19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.\*  
20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.\*  
21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.\*  
22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6377	100.0	1225	21	AA07973
2	6129	96.1	1247	21	AA07974
3	5633	88.3	1084	20	AAV23900
4	5426	85.1	1040	18	AAW24559
5	3916	61.4	756	19	AAW06664
6	2818.5	44.2	1192	20	AAI23899
7	2818.5	44.2	1211	18	AAW24560
8	2818.5	44.2	1211	21	AAV23336
9	833	13.1	837	21	AA036464
10	778.5	12.2	821	21	AA036467
11	770.5	12.1	797	21	AA036462

12	766	12.0	797	21	AA036466
13	719.5	11.3	726	21	AA036463
14	684	10.7	425	21	AA036469
15	256.5	4.0	1185	20	AAV33497
16	255.5	4.0	1012	20	AAV17406
17	222	3.5	802	19	AA037153
18	219.5	3.4	1837	21	AA035564
19	218.5	3.4	787	19	AA037152
20	218	3.4	733	19	AA037151
21	194.5	3.1	2842	15	AA03508
22	194	3.0	2843	16	AA011922
23	194	3.0	2843	19	AA076140
24	194	3.0	2843	19	AA076144
25	194	3.0	2843	21	AA035392
26	194	3.0	2973	19	AA076821
27	194	3.0	2973	19	AA076822
28	193.5	3.0	2385	21	AA085569
29	192	3.0	2843	13	AA026052
30	192	3.0	2843	18	AA035392
31	192	3.0	2843	19	AA035392
32	192	3.0	2860	15	AA035370
33	191.5	3.0	781	21	AA063507
34	191.5	3.0	781	21	AA070740
35	190	3.0	1099	21	AA070393
36	190	3.0	2973	21	AA070304
37	189	3.0	1099	17	AA051177
38	189	3.0	2971	21	AA041231
39	188.5	3.0	2518	21	AA040574
40	188	2.9	2972	22	AA050363
41	188	2.9	3118	22	AA050362
42	187	2.9	863	21	AA042952
43	187	2.9	1130	17	AA005178
44	187	2.9	1130	21	AA070390
45	187	2.9	1130	21	AA070394

#### ALIGNMENTS

RESULT 1	AA07973	standard; Protein; 1225 AA.
ID	AA07973	
XX	AA07973:	
XX	14-NOV-2000 (first entry)	
DE	A human neural plakophilin related armidillo protein.	
XX	Human; presenilin; PS1; PS2; nerve cell growth; nerve damage; hnpRap;	
KW	neural plakophilin related armidillo protein; Alzheimer's disease;	
KW	Parkinson's disease; amyotrophic lateral sclerosis; ALS; stroke;	
KW	multiple sclerosis; ischemic stroke; neural paropathy; sciatic crush;	
KW	motor neuron disease; peripheral neuropathy; neuropathy; diabetes;	
KW	spinal cord injury; facial nerve crush.	
XX		
OS	Homo sapiens.	
XX		
PN	W0200047615-A2.	
PD	17-AUG-2000.	
XX		
PF	11-FEB-2000; 2000MO-CA00126.	
XX		
PR	12-FEB-1999; 99US-0119835.	
XX		
PA	(UTOR) UNIV TORONTO GOVERNING COUNCIL.	
XX		
PI	St George-Hyslop PH, Fraser PE;	
DR	WPI: 2000-524531/47.	
XX	N-PSDB; AAA59700.	

Mouse plakophilin-  
Human plakophilin  
Human cancer assoc  
Human atrophin I p  
Human atrophin-1  
Mouse neutral Menat  
Human homologue of  
Mouse neutral Menat  
Adenomatous polyp  
Adenomatous polyp  
Human APC protein  
Human APC protein  
Human APC protein  
Transcriptional ac  
Human homologue of  
APC gene product i  
Human adenomatous  
Human adenomatous  
Adenomatous polyp  
Human beta catenin  
Human beta catenin  
Drosophila melanog  
Protein used in ca  
Lats large tumour  
Human OREX ORE95  
Human OREX ORE38  
Human SRAP. Homo  
Human SRAP. Homo  
Human OREX ORE2716  
H-Lats large tumou  
Human Lats (large

2000 data

Gene

PT Stimulation of nerve cell growth using human Neural Plakophilin Related  
 PT Armidillo Protein (hNRPAP) polypeptide, useful for the treatment of  
 PT diseases such as Alzheimer's, Parkinson's, and stroke  
 PS  
 XX  
 PS Claim 3: Page 24-27; 33pp; English.

CC The present sequence represents human Neural Plakophilin Related  
 CC Armidillo Protein (hNRPAP) polypeptide. hNRPAP interacts with presenilins  
 CC (PS1 and PS2) (PS1 and PS2). The specification describes a method  
 CC for stimulating the growth of nerve cells, comprising contacting  
 CC them with hNRPAP. The hNRPAP polypeptide and polynucleotide are useful  
 CC for treating nerve damage caused by a variety of diseases or physical  
 CC traumas, including Alzheimer's disease, Parkinson's disease, amyotrophic  
 CC lateral sclerosis (ALS), multiple sclerosis, stroke, ischemia associated  
 CC with stroke, neural paropathy, motor neuron diseases, sciatic crush,  
 CC peripheral neuropathy, neuropathy associated with diabetes, spinal cord  
 CC injuries and facial nerve crush.

CC Sequence 1225 AA:

Query Match 100.0%; Score 6377; DB 21; Length 1225;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 1225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MFARKPGAAPLGAAMPVDPDSASEKTSLSPLGNTSGDSEETTSALIAVKBOEL 60  
 DB 1 mfarpgpaaplgaampvdpdsasektslsplgntsgdsettsalialvkegel 60  
 QY 61 QFERLTRELEAKQIVASQLEKCKIGSETGSMSSMSAEQFOWOSODQDDELTG 120  
 DB 61 qferltreleaqivasaqlerckigsetgsmssmsaeqfowosodqddelittg 120  
 QY 121 LEIVNSCIRISQESGIIIDPOYSGERSSILSQSALQNSKPEGSFOYPAHYHNOITAL 180  
 DB 121 leivnsctrisqesgiiidpoystgerssilssqsalqnskegsfoypahynqital 180  
 QY 181 GETTPSQPLPARQARATGOSFSGTTSRAGHLAGEPAPRPPPPRPFAPSLGSFHL 240  
 DB 181 gettpsqplparqaratgosfsgttsraghlagpapppppprpfapslgsfhl 240  
 QY 241 PDAPAAAAALYSSSTLPAPRPGSGPLAPQSGSPKTLORGSGAPRGATYAAPRGSSP 300  
 DB 241 pdapaaaaalyssstlpaprrpgsgplapqsgspktlorgsaprgatyaaprgssp 300  
 QY 301 KQSPRLAKSVSTSPINIVWSAGLSPIRTVSPTVOSTISSSPIHOLSTIGYATLS 360  
 DB 301 kqsprrlaksystspiniwvssaglsprvtvsptvostissspiholstigtatls 360  
 QY 361 PTKRLVHASSEYSKHSELYATLQRPGLAAGSRASYSQHGHLGPELRALQSPENHI 420  
 DB 361 ptkrlvhasseyskhselyatlqrpglaaagsraaysqhghlgpelralqspenhi 420  
 QY 421 DPITYDRVYQKPPMSSLQSGDPLPRAHTGTTRTSPAPSPGVDSVPLQRTGSGOHQPN 480  
 DB 421 dpytdrvyqkppmsslsgsdplprahgttrtspapspgvdsvplqrtgsgohqpn 480  
 QY 481 AAAATFORASYAAGPASYADPYRQLOCPSESPYSKSGALPEGLASPSIDSTOK 540  
 DB 481 aaaatforasyaagpasyadpyrtqlycspespyssksgalpeglaspsidsstok 540  
 QY 541 DPREFGMRDPELEVTQMLQHQFSPVQSNAAAYLQHLFCGNKIKAEIRROGGIQLAVDL 600  
 DB 541 dprefgmrdepelvtqmlqhqfsvpvsnaaaylqhlfcgnkikaetrrroggiqlavdl 600  
 QY 601 LDHRTVEVHRSAGALRNLYVGKANDNKIALKNGGIPALVRLIKRTTDEIRLYTVGV 660  
 DB 601 ldhrtvevhrsagalnlyvgkandnkialknggipalvrlirkttdeirlytvgv 660  
 QY 661 LWNSSCDALMPTIQDMLAVITNAILPHSGWENSPLODRKIQIOLHSSQVLRNATGCLR 720  
 DB 661 lwnsscdalmptiqdmlavitnailphsgwensploidrkqiqlhssqvlrnatgclr 720

QY 721 NVSSAGEARRRRECDGLTDALLVYIOGALSGSEIDSKTVENCYCILRNLSYRLAETS 780  
 DB 721 nvssagearrrrrecdgltdallvyiogalsgesidsktvencycilrnlsyrlaets 780  
 QY 781 QGQHMGTDELIDGLGCEANGKDAESSGCWKKKKKKSSODQMDVGPLPDCAEPPKGIOM 840  
 DB 781 qgqhmgtdelidglgceangkdaessgcwkkkkkkssodqmdvgplpdcappkqiom 840  
 QY 841 LHMPSIVKRYTLTLSCSNPDLEGAAGLQNLAAAGSWKWSYITAAARKKEGFLITVEL 900  
 DB 841 lhmpsivkrytltlscsnpdleagaaglqnlaagswkwswyitaaarkekgflitvel 900  
 QY 901 LRIDNRVCAVATARNALDVNRNKLKGYAMRDVRLPLGNNNSNNTASKAMSDTV 960  
 DB 901 lrindrvcavatarnaldrvnrnklkgymrdvrlplggnnsnntaskamsdtv 960  
 QY 961 TAVCGTLHEVYTKNMENAKALRDAGIEKLVGISKSGDKHSPKYKAAQVLSNMOYR 1020  
 DB 961 tavcgtlhevtytkmenakalrdagieklygiskskgdkhspkykkaasqvlnsmoyr 1020  
 QY 1021 DLRSLYKKDQMSQYHFAVSSSTIERDRORPYSSRPTSPISPRVSPNNRSASAPSPREM 1080  
 DB 1021 dlrslykkdqwsgyhtvaassstierdrprysstrpsisprvspnnrsasapsprem 1080  
 QY 1081 ISLKERKTDYECTGSNATYHGAKGERTSRKDMTAQNTGISTLYNNSYGAPAEIDIKHNOY 1140  
 DB 1081 islkerktdeyctgsnatyhgakgertsrkdmtaqtngistlynnsgapaeidikhnoy 1140  
 QY 1141 SAQVPQEPSRKDYETVOPONSTRNWDSEFEDOVHNRPASETMHLGLKSTNYVDF 1200  
 DB 1141 saqvpqepsrkdyetvoponstrnwdsefedovhnrpasetmhlglkstnyvdf 1200  
 QY 1201 YSAARPYSELNETSHYPASPDQSW 1225  
 DB 1201 ysaarpyseletshypaspsdw 1225

# RESULT 2

AAB07974  
 ID AAB07974 standard; Protein; 1247 AA.

AC AAB07974;

DT 14-NOV-2000 (first entry)

DE A murine neural plakophilin related armidillo protein.

KW Human; presenilin; PS1; PS2; nerve cell growth; nerve damage; hNRPAP;

KW neural plakophilin related armidillo protein; Alzheimer's disease;

KW Parkinson's disease; amyotrophic lateral sclerosis; ALS; stroke;

KW multiple sclerosis; ischemia; stroke; neural paropathy; sciatic crush;

KW motor neuron disease; peripheral neuropathy; neuropathy; diabetes;

KW spinal cord injury; facial nerve crush.

OS Mus sp.

PN W0200047615-A2.

PD 17-AUG-2000.

PF 11-FEB-2000; 2000MO-CA00126.

PR 12-FEB-1999; 99US-0119835.

PA (UTOR ) UNIV TORONTO GOVERNING COUNCIL.

PI St George-Hyslop PH, Fraser PE;

DR WPI; 2000-524531/47.

XX N-PSDB; AAA59701.

PT Stimulation of nerve cell growth using human Neural Plakophilin Related  
 PT Armidillo Protein (hNRPAP) polypeptide, useful for the treatment of

PT diseases such as Alzheimer's, Parkinson's, and stroke -

XX PS Disclosure: Page 30-33; 33pp; English.

XX The present sequence represents Neural Plakophilin Related Armadillo Protein (NPRAP) polypeptide. Human NPRAP interacts with presenilins (PS) I and II (PS1 and PS2). The specification describes a method for stimulating the growth of nerve cells, comprising contacting them with hNPRAP. The hNPRAP polypeptide and polynucleotide are useful for treating nerve damage caused by a variety of diseases or physical traumas, including Alzheimer's disease, Parkinson's disease, amyotrophic lateral sclerosis (ALS), multiple sclerosis, stroke, ischemia associated with stroke, neural paropathy, motor neuron diseases, sciatic crush, peripheral neuropathy, neuropathy associated with diabetes, spinal cord injuries and facial nerve crush.

XX Sequence 1247 AA:

Query Match 96.1%; Score 6129; DB 21; Length 1247;  
Best Local Similarity 94.8%; Pred. No. 0;  
Matches 1185; Conservative 8; Mismatches 29; Indels 28; Gaps 3;

DB 1 MFARPGAGAPLGAAPVDPDSSASEKTSLSPLNTSNGDSEETTSAILASVKEQEL 60  
1 mfarpgagaplgampypdqpssaseknsslsplngdsettsailasvkeqel 60

DB 61 QFERLTRELEAEROTIVASQLEKCKLGESETGSMSSAEEQFOQSDGOKDDELTTC 120  
61 qferltreleaeerivasqlercklgetsgsmssaaeeqfoqsdgokdedelttc 120

DB 121 LEIVDSCTRSIQESGILDPDYSSTGERPSLSQSALQLSNKPESFOYAPASVSNQTLAL 180  
121 leivdsctrsiqesgilddpdyssstgerpsllsqsalqlsnkpesfoypasvsnqtlal 180

DB 181 GETTPSQLPARGTOARATGSEFSGTSTRAGHLAGPEPAPPPPPPPPPPPPPPPPPPP 240  
181 gettpsqlpargtoaratgsefsgtstraghlagpepapppppppppppppppppppp 240

DB 181 gdtapsqparstqaraagsfsgtctgtaghlagsepa-pppppppppppppppppppp 239

DB 241 PDAPPAALAAALYSSSTPLAPPPGSGPLAAPGSGPTKLRGSGAPPEATYAAAPRGSSP 300  
241 pdapp-aaalysstplappgsgpllaapgsgptklrgsgappeatyaaaprgssp 300

DB 298 KQSPRLAKSYSTSPINIVSSAGLSPTRVSPPTVOSTISSPTIHQSSTIGTYATLS 360  
298 kqsprlaksystspinivssaglsptrvspptvostissptihqsstigttyatls 360

DB 361 PTKRLVHASBOYSKHQELATATLQRPCSLAAGSRASVSGHGLGPELRALQSPENHI 420  
361 ptkrlvhasboyskhqelatatlqrpcslaagsrasvsgghlgpeleralqspenhi 420

DB 358 ptkrlvhasseqkhsqelatatlqrpcslaagsrasvsgghlgpeleralqspenhi 417

DB 421 DPTIERYVYQKPKMRSLSSQSGDPLPRAHTGYRTSTABSPGVSVPLORTSGHGPON 480  
421 dptieryvyqkpkmrslssosgdpplrphatgyrtstabspgvsvploatsghgp 480

DB 418 dptiedrvyqkpmrslsqsgdplrpahltgtrstabspgvsvploatsgshgpn 477

DB 481 AAAATFORASVYAGPANTADPYROLOYSPVESPYSKSGAPLPEEGTLASPSIDSIQK 540  
481 aaaatforasvyaagpantadpyroloyspvespyksagplpeegtlaspsidsiqk 540

DB 478 aaaatfgrasvyaagpantadpyrtdycaavdspsysgpalppegtlarspsidsiqk 537

DB 541 DPREGWRPELPEVLOLQHPFSVQNAALVLOHLCFGDNKIKAEIRROGGIQLVLDL 600  
541 dpregwrpelpevlolqhpfsvqnaalvlohlcfgdnkikaetrrroggiqlvldl 600

DB 538 dprefgrwpelepevlqmqhfpvgsnaaaylqhlfcgdnkikaetrrr99qqlvldl 597

DB 601 LDHRTVEHRSAGALRNLYVYKANDDKIALKNCGGITPALVRLIKRTTDLIRELVTGV 660  
601 ldhrtvehrsagalnlyvykanddkialkncggitpalvrlirkrttdlirelvtgv 660

DB 598 ldhrtvehrsecgalmlyvykanddkialkncgipalvrlirktdleirelvtgv 657

DB 661 LWNLSGCDALKMPITODALAVLTNAVLIIPHSGWENSPLODDRKIOHLSOVLRAATGCLR 720  
661 lwnlsgcdalkmpitodalavltnavliiphsgwen脾oddrkiohlssovlraatgclr 720

DB 658 lwnlsgcdalkmpitodalavltnavliiphsgwen脾oddrkiohlssovlraatgclr 717

DB 721 NVSSAGEARRMRRCGDTLADLVVIOALGSSEIDSKTVENCVCILRNLSYLAETS 780  
721 nvssagearrmrrecgdtladlvvialgsseidsktvencvcilrnlsylaets 780

DB 718 NVSSAGEARRMRRCGDTLADLVVIOALGSSEIDSKTVENCVCILRNLSYLAETS 777

QY 781 OGGHMTDELDELGCGEANGKDAESSGCMKKKKKKKKKKKKKKKKKKKKKKKKKKKK 840  
781 ogghmtdeldelgcgeangkdaessgcmkkkkkkkkkkkkkkkkkkkkkkkkkkkk 840

DB 778 qggmgtdelddlgelcngkdaessgcmkkkkkkkkkkkkkkkkkkkkkkkkkkkk 837  
778 qggmgtdelddlgelcngkdaessgcmkkkkkkkkkkkkkkkkkkkkkkkkkkkk 837

QY 841 LHMPSLVKRYLTLLSECSNPDTLEGAGALQNLAAAGSWK----- 879  
841 lhmplsvkryltllsecsnpdtlegagalnlaagswk----- 879

DB 838 lhmplsvkryltllsecsnpdtlegagalnlaagswk----- 879  
838 lhmplsvkryltllsecsnpdtlegagalnlaagswk----- 879

QY 880 ----SVYRAAVREKELPLVELLRIDNRVVCAYATLRNALDVNKKELIGYAMR 935  
880 ----svyraavrekelpvelrlidnrvcayatlrmaladvnkkeligyamr 935

DB 898 clpqsavviraavrekelpvelrlidnrvcayatlrmaladvnkkeligyamr 957  
898 clpqsavviraavrekelpvelrlidnrvcayatlrmaladvnkkeligyamr 957

QY 936 DLVHRLPGGNNMNTASKMSDDVTAVCCVLEHVRKNNENKALPDAGITKLVGISK 995  
936 dlvrhrlpggnnmntasksmsddvtavccvlehvrrknnenkalpdagitklvgisk 995

DB 958 dlvrhrlpggnnmntasksmsddvtavccvlehvrrknnenkalpdagitklvgisk 1017  
958 dlvrhrlpggnnmntasksmsddvtavccvlehvrrknnenkalpdagitklvgisk 1017

QY 996 SKGDHRSKRYKAAQVLNSWQYRDLRLSLYKKGWSQYHFAVSSSTIERDORPYSSSR 1055  
996 skgdhrrskrykkaaqlvlnswqyrdlrlsllykkgwsqyhfavssstierdorpys 1055

DB 1018 skgdhrrskrykkaaqlvlnswqyrdlrlsllykkgwsqyhfavssstierdorpys 1077  
1018 skgdhrrskrykkaaqlvlnswqyrdlrlsllykkgwsqyhfavssstierdorpys 1077

QY 1056 TFSISPVFVSPNNNSASAPASPREMISLKERKDYECTGSNATYHGAKGPHSRKDAMTA 1115  
1056 tfsispvfvsppnnnsasapaspremislkerkdyectgsnatyhgakgphsrkdamta 1115

DB 1078 tpsispvfvsppnnnsasapaspremislkerkdyectgsnatyhgakgphsrkdamta 1137  
1078 tpsispvfvsppnnnsasapaspremislkerkdyectgsnatyhgakgphsrkdamta 1137

QY 1116 QNTGISLTYRNSYGAPEADIKHNOVASQVPOPSRKYDTFTYPOFNSTRNDESEFED 1175  
1116 qntgisltyrnsygapadikhnovasqvpopsrkydfttypofnstrndesefed 1175

DB 1138 qntgisltyrnsygapadikhnovasqvpopsrkydfttypofnstrndesefed 1197  
1138 qntgisltyrnsygapadikhnovasqvpopsrkydfttypofnstrndesefed 1197

QY 1176 VHHRRPASEYTMHGLKSTGNYDFYSAAAPYSELNFTSHYPASDPSW 1225  
1176 vhhrrpaseytmhglkstgnydfysaaapyseelnftshypasdpdsw 1225

DB 1198 vhhrrpaseytmhglkstgnydfysaaapyseelnftshypasdpdsw 1247  
1198 vhhrrpaseytmhglkstgnydfysaaapyseelnftshypasdpdsw 1247

RESULT 3  
AAV23900  
ID AAV23900 standard; protein; 1084 AA.  
XX  
AC AAV23900;  
XX  
DT 27-SEP-1999 (first entry)  
XX  
DE Human resenilin binding armadillo protein GT24/hNPRAP.  
XX  
KW Human; presenilin 1; PS1; presenilin-binding protein; interacting domain;  
KW presenilin allele; Alzheimer's disease; senile dementia;  
KW psychiatric disease; schizophrenia; depression; neurological disease;  
KW stroke; cerebral haemorrhage; p0071; armadillo protein.  
XX  
OS Homo sapiens.  
XX  
FH  
FT Key Location/Qualifiers  
FT Misc-difference 115 /note= "not specified"  
XX  
PN MO9935501-A1.  
XX  
PD 15-JUL-1999.  
XX  
PF 08-JAN-1999; 99WO-CA00018.  
XX  
PR 09-JAN-1998; 98US-0070948.  
XX  
PA (UTOR ) UNIV TORONTO GOVERNING COUNCIL.  
XX  
PI Fraser PE, St George-Hyslop PH;  
XX  
DR WPI, 1999-419410/35.  
XX  
PT Identifying substances that alter presenilin interactions, useful  
PT for screening individuals for presenilin alleles associated with  
PT Alzheimer's disease - useful for diagnosis of Alzheimer's disease

XX Claim 9; Fig 1; 40pp; English.  
 PS The present sequence represents a human presenilin-binding protein,  
 CC termed GR24/human neuronal plakoglobin related armadillo protein (GR24/  
 CC hNRPAP). The specification describes a method for identifying substances  
 CC that alter the interaction of a presenilin with a presenilin-binding  
 CC protein. The method comprises contacting the interacting domain of a  
 CC presenilin protein to a presenilin-binding protein in the presence of  
 CC a test substance, and measuring the interaction of the presenilin and  
 CC the presenilin-binding protein. The method can be used to screen  
 CC individuals for presenilin alleles associated with Alzheimer's disease  
 CC and related disorders, such as senile dementia's, psychiatric diseases  
 CC such as schizophrenia and depression, and neurological disease, such as  
 CC stroke and cerebral haemorrhage.  
 CC  
 CC Sequence 1084 AA;  
 SQ  
 Query Match 88.3%; Score 5633; DB 20; Length 1084;  
 Best Local Similarity 99.8%; Pred. No. 0;  
 Matches 1076; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 148 PSLTSSQATLQNSKPEGSFOYPASYSNOTLALGETTPEQLPARGTOARATGSEFSOGTT 207  
 7 PSLTSSQATLQNSKPEGSFOYPASYSNOTLALGETTPEQLPARGTOARATGSEFSOGTT 66  
 208 SRAGHLAGEPPAPPP 267  
 67 SRAGHLAGEPPAPPP 126  
 268 PLAAPGGSTKIQRGSAPEGATYAAPRGSSPKSPSLAKSYSPSSINIVYSSAGLS 327  
 127 PLAAPGGSTKIQRGSAPEGATYAAPRGSSPKSPSLAKSYSPSSINIVYSSAGLS 186  
 328 PIRVTSPTVQSTISSPSIHQLSTITGTYATLSPTKRLVHASEQYSKHSOELATITLDR 387  
 187 PIRVTSPTVQSTISSPSIHQLSTITGTYATLSPTKRLVHASEQYSKHSOELATITLDR 246  
 388 PGSLAAGSRASYSQSGHGLPELRLAQLSPHHIDTYEDVYQKPPMRSLISQSGDPLPP 447  
 247 PGSLAAGSRASYSQSGHGLPELRLAQLSPHHIDTYEDVYQKPPMRSLISQSGDPLPP 306  
 448 AHNTGTYRTAPSPGVDVYPLQRTGSOHGPONAAATFORASYAAGPASNVADPYRLQ 507  
 307 AHNTGTYRTAPSPGVDVYPLQRTGSOHGPONAAATFORASYAAGPASNVADPYRLQ 366  
 508 YCPSPESPYKSGPALPPEGLTARSPSIDSIQKDPREFGMRDEPLPVYIOMLOHOPSVQ 567  
 367 YCPSPESPYKSGPALPPEGLTARSPSIDSIQKDPREFGMRDEPLPVYIOMLOHOPSVQ 426  
 568 SNAAYIQLHCFQGNKTKAEIRROGIGIOLVDLDDHMTVEYHRSAGALNRLVYKANDD 627  
 427 SNAAYIQLHCFQGNKTKAEIRROGIGIOLVDLDDHMTVEYHRSAGALNRLVYKANDD 486  
 628 NKIALKNGGIPALVRLKRTTDEIRLTVGVLMNLSGCDALMPITODALAVLTNAVY 687  
 487 NKIALKNGGIPALVRLKRTTDEIRLTVGVLMNLSGCDALMPITODALAVLTNAVY 546  
 688 IPHSGWENSPLODDRKTIOHSSOYLKRNATGCLRWVSSAGFEARRRRCOGITALLVYI 747  
 547 IPHSGWENSPLODDRKTIOHSSOYLKRNATGCLRWVSSAGFEARRRRCOGITALLVYI 606  
 748 OSALGSSRIDKTYENCVCILNLSYRLAETSOGHGTDELGLLGEANGKDAESSG 807  
 607 OSALGSSRIDKTYENCVCILNLSYRLAETSOGHGTDELGLLGEANGKDAESSG 666  
 808 CWGKK 867  
 667 CWGKK 726  
 868 GALQNLAAAGSMKWSYITRAAVRKEKGLPIVVELLRINDRVYCAVATLRNNALDVRNKE 927

DB 727 galqnlaaagsmkwsyitraavrkekgplivvelrlindrvycaavatalrnnaldvrnke 786  
 QY 928 LICKYAMRDLVHRLPBGNNNSNTASKAMSDPTVAVCCITHEVITKNNENAKALRDAGCI 987  
 DB 787 LICKYAMRDLVHRLPBGNNNSNTASKAMSDPTVAVCCITHEVITKNNENAKALRDAGCI 846  
 QY 988 EKIYVIGISKSGDKHSPKVVYKASQVYLSNMWYRDLSLYKKDQMSQYHFVASSSTIENDR 1047  
 DB 847 EKIYVIGISKSGDKHSPKVVYKASQVYLSNMWYRDLSLYKKDQMSQYHFVASSSTIENDR 906  
 QY 1048 QRPYSSRTSPISIPVYSPRNRSASAPSPREMLSKERTYECTGNSNATYHAKGEHT 1107  
 DB 907 QRPYSSRTSPISIPVYSPRNRSASAPSPREMLSKERTYECTGNSNATYHAKGEHT 966  
 QY 1108 SRKQAMTAQNTGISTLYRNSYGAPADIRKNOVSAQVPQEDSRKDYETTYQPNSTNRY 1167  
 DB 967 SRKQAMTAQNTGISTLYRNSYGAPADIRKNOVSAQVPQEDSRKDYETTYQPNSTNRY 1026  
 QY 1168 DESFEEDQVHHRPPASITVHGLKSTGNYVDYSAARPSSELYETSHYPAASPSMV 1225  
 DB 1027 DESFEEDQVHHRPPASITVHGLKSTGNYVDYSAARPSSELYETSHYPAASPSMV 1084  
 RESULT 4  
 AAW24559 standard; Protein: 1040 AA.  
 ID AAW24559;  
 AC AAW24559;  
 XX 06-FEB-1998 (first entry)  
 DE Presenilin-Interacting protein GT24.  
 KW Presenilin-Interacting protein; human; Alzheimer's disease;  
 KW diagnosis; therapy; transgenic animal; animal model; GT24.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT Domain 346..862  
 FT /note= "presenilin-interacting domain"  
 FT Misc-difference 12  
 FT /note= "encoded by GST"  
 FT Misc-difference 40  
 FT /note= "encoded by CCR"  
 FT Misc-difference 71  
 FT /note= "encoded by AKC"  
 XX  
 PN WO9727296-A1.  
 PD 31-JUL-1997.  
 XX  
 PF 27-JAN-1997; 97WO-CA00051.  
 XX  
 PR 02-JAN-1997; 97US-0034590.  
 PR 26-JAN-1996; 96US-0592541.  
 PR 05-JUL-1996; 96US-0021673.  
 PR 12-JUL-1996; 96US-0021700.  
 PR 08-NOV-1996; 96US-0029895.  
 XX  
 PA (HSCR-) HSC RES & DEV LP.  
 PA (UTOR ) UNIV TORONTO GOVERNING COUNCIL.  
 PI Frazer PE, Rommens JM, St George-Hyslop PH;  
 DR N-PSDB; AAT79967.  
 XX  
 PT Presenilin-interacting protein genes - used to develop products for  
 PT the diagnosis, therapy and study of Alzheimer's disease and related  
 PT disorders  
 PS Claim 1; Page 99-101; 133pp; English.





CC arm repeat that is characteristic of catenin proteins. Overall,  
 CC ALARM and delta-catenin are 52.3% similar and 32.1% identical. The  
 CC invention provides isolated ALARM polypeptide, nucleic acids,  
 CC vectors, host cells and specific antibodies. Presentin 1 protein  
 CC can be detected in samples (especially human cerebrospinal fluid),  
 CC by contacting with ALARM (claimed). Mutation of the presentin 1  
 CC gene is associated with familial Alzheimer's disease. ALARM can be  
 CC used to diagnose human diseases caused by a mutant form of a protein  
 CC interacting with ALARM, by analysing fluid samples to detect the  
 CC ALARM-interacting protein (claimed). ALARM can also be used to  
 CC diagnose altered levels of presentin 1, by contacting a sample with  
 CC ALARM and determining binding (claimed). ALARM polypeptides and  
 CC antibodies can be used to detect an ALARM-containing complex in  
 CC biological samples, by contacting with polypeptide or antibody and  
 CC determining whether the polypeptide/antibody binds to a sample  
 CC component (claimed). The antibodies are also useful to detect ALARM  
 CC polypeptides and to measure the effect of candidate compounds on  
 CC expression or localisation of ALARM. They are useful to evaluate  
 CC engineered cells prior to introduction in gene therapy, to inhibit  
 CC abnormal ALARM activity or to generate anti-idiotypic antibodies,  
 CC useful therapeutically.

Sequence 756 AA:

Query Match 61.4%; Score 3916; DB 19; Length 756;  
 Best Local Similarity 94.9%; Pred. No. 8.1e-272;  
 Matches 752; Conservative 0; Mismatches 4; Indels 36; Gaps 1;

QY 434 MRSLSQSGDPLPAHNTYRTSTAPSSPGVDSPVLORTSGHGPONNAATFCRASYAA 493  
 DB 1 mrslsqsgdplpahntyrtstapsspgvdspvldlrdhmrtevhhsac 60  
 QY 494 GPASNYADPYRLOLCPESVPSKSGPALPPEGTLASPISIDSKDPPREGRDDEL 553  
 DB 61 gpasnyadpyrloqcpesvpsksgpalppegtlarspsidskdprefgrwdep 120  
 QY 554 EVIOMLQHPFVSVOANAAVLQHLCEFGDNKIKAEIRROGGIQLLVLDLDRHRTVHRSAC 613  
 DB 121 eviomlqhpfvsvoanaavlyqlhlcfgdnkikaelrrggqllvldldhmrtevhhsac 180  
 QY 614 GARNLYVYGKANDNKIKAKCGGIPALVRLRKTDTLELTVGYLWNLSSCDALKMP 673  
 DB 181 garnlyvygkandnkialkncggipalvrlrktdtlelvtgylwmlsscdalkmp 240  
 QY 674 ITDPAALVLTAAVLIIPHSWENSPLODDRKTQLHSSQVLRNATGCLNNVSASGEAHRM 733  
 DB 241 itdpaalvltavliiphsweensploddrkqlhssqylmatgclnnvsasgeahtmm 300  
 QY 734 RECGLDLALYVIOALGSSEIDSKTVENCVCILRLSYRLAETSQGHMTDEL 793  
 DB 301 recgldlallvyioalgsseidsktevencvclrlsylaetsqghmtdeidgl 360  
 QY 794 LCGEANGDASSGCGMKKKKKKKSSODOMOVGVLPCAPKPKIOMLWHPSTYKPVLT 853  
 DB 361 lcgeargdaasscgwkkkkkkksqdgwvlpdcaepkyiqnlwpsvlykpyltl 420  
 QY 854 LSECSNPDTLEGAAGALONTLAAGSMKWSVYTRAVERKEGPIIVELRLRIDNRRVCAVA 913  
 DB 421 lsecsnpdtlegaaqalnlaagswkvsylraavtrkekprlivelrlridnrvacava 480  
 QY 914 TALNNMALDVNRKELIKYAMRDVHRLPGGNSNNTASTRAMSDDTYAVACCTLHEVITK 973  
 DB 481 talnnmaldvnrkelikyamrdlvhr1pggnsnntastkamsddltavacctlhevltk 540  
 QY 974 NMENAKALRDAGGIEKLVGTSKSGDKHSPKVVKAASOVNMSMVOYDLRLSKKKQMSQ 1033  
 DB 541 nmenakalrdaggiekvlgtsksgdkhspvkaasgvlnsmqyrdlsllykkkgwsq 600  
 QY 1034 YHFVASSSTIERDRORPYSRRTPSPISPVRYSPNNRSASAPASPREMISLKEKRTYECP 1093  
 DB 601 yhfvaassstierdrorpyssrrtpspisprvrsppnnsasapasprenmislkerktdelect 660

QY 1094 GSNATYGAKEGHTSRKADAMTAONTGISTLYRNSYGAPADIKHNOVSAQPVQPEPRKD 1153  
 DB 661 gsnatygakeghtsrkdamttaontgstlyrnsygapadikhnovsaqpvqpeprkd 692  
 QY 1154 YETYPQNSTRYNDSFEEDQVHRRPPASEYTMHLGKSTGNYDYFYSARPYSELNTE 1213  
 DB 693 -----nstrydesffedqvhrrppaseytmhlgkstgnydyfysarpysele 744  
 QY 1214 TSHYPASPSDW 1225  
 DB 745 tshypaspsdw 756

RESULT 6  
 ID AAY23899 standard; protein; 1192 AA.  
 AC AAY23899;  
 XX 27-SEP-1999 (first entry)  
 DT XX  
 DE Human resenilin binding armadillo protein p0071.

XX OS Homo sapiens.  
 XX PN W09935501-A1.  
 XX PD 15-JUL-1999.  
 XX PF 08-JAN-1999; 99WO-CA00018.  
 XX PR 09-JAN-1998; 98US-0070948.  
 XX PA (UTOR ) UNIV TORONTO GOVERNING COUNCIL.  
 XX PI Frazer PE, St George-Hyslop PH;  
 XX DR WPI; 1999-419410/35.  
 XX XX

PT Identifying substances that alter presentin interactions, useful  
 PT for screening individuals for presentin alleles associated with  
 PT Alzheimer's disease - useful for diagnosis of Alzheimer's disease  
 XX  
 PS Claim 10; Fig 1; 40pp; English.

XX The present sequence represents a human presentin-binding protein,  
 CC termed p0071. It is a member of the armadillo family of proteins.  
 CC The specification describes a method for identifying substances  
 CC that alter the interaction of a presentin with a presentin-binding  
 CC protein. The method comprises contacting the interacting domain of a  
 CC presentin protein to a presentin-binding protein in the presence of  
 CC a test substance, and measuring the interaction of the presentin and  
 CC the presentin-binding protein. The method can be used to screen  
 CC individuals for presentin alleles associated with Alzheimer's disease  
 CC and related disorders, such as senile dementia's, psychiatric diseases  
 CC such as schizophrenia and depression, and neurological disease, such as  
 CC stroke and cerebral haemorrhage.

Sequence 1192 AA:

Query Match 44.2%; Score 2818.5; DB 20; Length 1192;  
 Best Local Similarity 49.2%; Pred. No. 6.6e-193;  
 Matches 618; Conservative 192; Mismatches 337; Indels 109; Gaps 37;

QY 15 MWPVQPPSSASKTSLSBGLNWSMGDSFETET-SALLASVKROELOFELTRELAEKR 73  
 DB 1 mpvqpssasaktslsbglntsmgdsfett-sallavkroelofeltrelaelevr 59



	Matches	618:	Conservative	192:	Mismatches	337:	Indels	109:	Gaps	37:
QY	15	MPVPPQSSASAKETSSLSGLMTNSDGEETTT--SALLASVKEQELQRELRITRELAER	73							
Db	1	mpapeqaslvveeqpqltreqa--astipgmepetlalttllasvkeqelqrltrelevter	59							
QY	74	QIVASOLEKCKIGSEFGSSMSMSAEEQOMQOD-----GOKDEDELTGLELVDSOC	127							
Db	60	qlvasqletcrcligaeeprslaststestksfpwsttdvtrpntgvskprvdsavqpmny1----	115							
QY	128	ISLSDESGTL-DRQDTSTGERPSFL--SOSALQIOLNSKPEGSFOYRASHYNSQTLA----	179							
Db	116	lrltepegdgllysepeqfslhesegslgnrststqmsydsdgygeagsflnsgvskadnr	175							
QY	180	-----LGETTPRSQDLRAGTQARATGQSGFSQ--GTTSSAGHLADREARPRP-----	224							
Db	176	qghsfstgstinhhv-----instaeqqlvqpsvaniamtrvgsavpsraagspsvystgvs	231							
QY	225	PREPREPAPSLGSAFHLR--DAPRAAAAALYYSSSTTPRAPRPGSPPLAAPGGSPTKIQ	281							
Db	232	pergslrltalgsgfsgpsvtdrprlmpsa---yssttlpla--areaspy--sqpaasptair	286							
QY	282	RGGSAPREGATYAARPGSSPK--OSPSPILAKSYSTSSPINITVSSAGISPTIRTSPTVQST	340							
Db	287	tlgvsrlstsqtl--snuprprpqrqtatav-----gsplltldatq-----rvaspgqgq-v	333							
QY	341	ISSSPILHOLSTTIGTYATLSPT--KRLVNASQOYSKSHSOELATATLQRPGLSALASRASRY	399							
Db	334	gassp--krsgmtavqrhlqprslqrltvlhmeqfsgqyqydlgetwmprrpds1-tglrsay	390							
QY	400	SSOHNLGRELRLALDSPENHIDRTIEDRYQOKRPMKSLSSQSGDRLPRMNTGTTRTSTRAP	459							
Db	391	asqhnqqlqgdrlrsavspdlhltrlyegrtfuyasvpyrshpbtve--lqgsqetalyt----	445							
QY	460	SSPGVDVSV-PLQRTSSOHQONAAATQCRASYAAGPASNVDPRKQLOLQCSVSPYSK	518							
Db	446	---gysrdqgnldtasqr-----stltygdmlyntaltatqexprlqyrtqrq--ecynnr	496							
QY	519	SSPRLPREGTLARSSIDSIOKDPREFGKRDELREVIOMLOHQRPSVQSNAAATLOHLC	578							
Db	497	lqhavpaddgtlrrpslslslqkprfetaqrdeprvhlmlhqbfpvqanaaa1qlhlc	556							
QY	579	FEDNKTKAIEIRQOGLQCLVDLDLHRMTVEVNHSSAGCALRNLYUGKAGKANDPKALNCNGGI	638							
Db	557	fgdnkvwmevcrlggikhvldlndrvylevqnaagallnlyfsgstdeanklanknvg91	616							
QY	639	PALVRLKRTTDLERELVTVGLVNLMSODALKMPITODALAVLTNAVITPHSGWENSPL	698							
Db	617	palllrlfksldaeyretvltvgvltwlsocdawkmtllldalstlntvlypbgwnssff	676							
QY	699	ODDRKIOIHLSSOVLENNATGCLRNVSAGCEARRMKREDDGGLDALLXYLOSLGSSSEDS	758							
Db	677	dddhkikfqtslvtnnttgcrlnltsaqeaaekqmrsoeglyvdslllylthctvnstsyds	736							
QY	759	KTVENCVCILRNLSYRLAETSQOGHMGTDELDGLLCEBANGKDAESSGCGWKKKKKKK--	817							
Db	737	ktvencvcitlnlsyrltelyepqrartllglneldllgkespskdeps--cwjkkkkkkkr	795							
QY	818	---SODMKGVGRPLRCAPARRKGIOMLNMPSIYKRVLTLLSECSNPDTEGAAGALQNLAA	875							
Db	796	tpqedqwdqavprlprlsasprkxvemiwhpsvvrkrylltllaesnpaltleagsags1qlnlsa	855							
QY	876	GSMSKSVYTRAAVRREKELPILVELLRIDNDNRVYCAVNTALRNMLADVRNKEELIKUYMR	935							
Db	856	snwfaeaylvggprkrlprllvcllnndrvtvssgsatlmalmatdrtnkel1lkykmr	915							
QY	936	DLVHRLPEGNNSNMTASKAMSDDTVTAVCCTLHEVITKMMENAKALRDAGIEKLVGISK	995							
Db	916	dlvnr1prgngps-----vlsdetmaacalchevtskmenakaladsgglek1vnltk	970							
QY	996	SKGDKHSKVKVKAASOVLYNSMWQYRDLASLKQKQSQOYHNVAASSSTIERORQRYSSSR	1055							
Db	971	grtdasrltkvkaaqvntlclwgyrdlrslykkqgwqnhltpvstlelrdfksh----	1026							

OY		1066	TFSPISPV--RVSPPNNRSASAAAPREMI <del>S</del> LKERKTDYCCTGSMTNHCAGKEHTSRKDAM	1113
Dd		1027	-pslttngqmpslqsgvstgsscspallglrtidprseyrtlrpmqyynsgsdalhk--gl	1083
OY		1114	TNAONTGISTLTINSGCAPAE----IKNOVSAQDPVEPSPRKDYEYOPFONSTRNYDE	1169
Dd		1084	ypgsasvpshrppasaytmhlgshtgnvddrysaarpyseelnertshypaspdswnv	1125
OY		1170	SFEFGCVHHRPPASAYTMHLGSHTGNVDDRYSAARPYSEELNETSHYPASPDSWNV	1225
Dd		1142	pfyddrvhf-pastcytsctgyqlkstnlyvdylstkrr-----syraegyyspsdsww	1192
 RESULT       8 ID             AAY92336 standard; Protein: 1211 AA. XX AC             AAY92336; XX DT             21-AUG-2000 (first entry) XX DE             Human p0071. XX KW             NLK1 interacting protein; p0071; protein complex; cytosstatic; antiviral, KM             neuroprotective; cardiant. XX OS             Homo sapiens. XX PN             WO200020448-A2. XX PD             13-Apr-2000. XX PF             06-OCT-1999;      99W0-US23314. PR             06-OCT-1998;      98US-0167206. XX PA             (CURA-) CURAGEN CORP. PI             Nandabalan K, Schulz VP, Yang M; DR             MPI: 2000-303742/26. XX PS             N-FSDB; AAA09308.				
Pt			New complex of a NLK1 protein and a NLK1 protein-interacting protein,	
Pt			useful for treating cancer, hyperproliferative disorder,	
Pt			neurodegenerative disorder, cardiomyopathies, viral infections and	
Pt			metabolic disorders	
 Example 1: Page 154-159; 172pp; English.				
xx		AAY92331-37	were isolated in a modified yeast two hybrid system using	
cc		NLK1 protein as "bait".	These are known sequences which are Nlk1	
cc		interacting proteins.	The invention concerns purified complexes of a	
cc		NLK1 protein and a NLK1 protein-interacting protein,	where the	
cc		interacting protein is chosen from TrkA, protein phosphatase alpha,		
cc		14-3-3epsilon, alpha-tropomyosin, vimentin, p0071, Inl-1, IP-1, IP-2,		
cc		IP-3, IP-4, or IP-5.	Nlk1 (also referred to as Nek2) is a human	
cc		homologue of the Aspergillus nidulans mitotic regulator, NIMA kinase.		
cc		Nlk1 is a serine/threonine-specific kinase and is thought to play a key		
cc		role in cell-cycle events leading to the onset of mitosis.	The	
cc		complexes, their derivatives and Nlk1 or Nlk1-IP protein and DNA		
cc		sequences, etc. are useful for treating or preventing a disease or		
cc		disorder involving aberrant levels of the complex or protein.	Such	
cc		disorders include cancer, hyperproliferative disorders,		
cc		neurodegenerative disorders, cardiomyopathies, viral infections and		
cc		metabolic disorders.		
xx				
XO		Sequence      1211 AA;		

Query Match          44.2%; Score 2818.5; DB 21; Length 1211;

Best Local Similarity   49.2%; Pred No. 6.8e-193;

Matches 618; Conservative 192; Mismatches 337; Indels 109; Gaps 37;

```

QY 15 MPVPDPSSASKESTSLSPGLNTSGNDGSETET--SALIASVKEOELOFERLELEAER 73
Db 1 mpapeasaveeqpqtqrea-astcpamepetatcllaavkeqefqqltralever 59
QY 74 QIVASOLEKCKLGSETGSMSSMSAEQPOWQOD-----GQKDEDELTTGLELVDSO 127
Db 60 qivaeglerclrgaesplastesteksfwrstcvntgvekprvsdavqpnnyl----- 115
QY 128 INSLQESGL-DBDYSTGERPSL--SQAQLQNSKPEGSQYRASHSNQTLA----- 179
Db 116 itrepegglypseglleesegslgnsrscqmsysdsygeagsfmsqvsakadr 175
QY 180 -----LGETPSPQLPARAGTQARATGQSFQ--GTSRAGHLAEPAPPP----- 224
Db 176 qghsfistgnhv-----trstaegqlvgsvantrmrsvsyragspyvstgvs 231
QY 225 PPREPAPSLGSAFHLR---DAPPAALALYSSSTLPAPRGGSPLAAPGSGSPKIQ 281
Db 232 pargstlsgsfgspsvtdprlmpsa---ysstllpa-araaspy-sgrpasptair 286
QY 282 RGSABEGATYAPRGSSPK--QSPSLASTYSTSPINIVWSAGLSPTRTVSTVQST 340
Db 287 llysvtsrql-snpqprlpqyqltar-----gsplltldatq-----rvspsgqg-v 333
QY 341 ISSPPIHQLSTRTGATATLPT-KRLVHASQYSKHSCQELVATLQRPESLAAGSRASY 399
Db 334 gsssp--krsqmatavpqlhgsalqtrvhmeqfsgqgqyilyemvpprpsl-tglrsy 390
QY 400 SSOHGLPELRLALOSPENHIDPIYEDRYVQKPMRSLQSCQDPLPRAHTGTRSTAP 459
Db 391 aeqhsqldlrsavspdlhlprlyegrcyspvsrphgve-lgsqqlalyt----- 445
QY 460 SSGGVDSV-PLQRTGSOHQONNAATFORASYAAGPASNADYQQLCYCESVESPEY 518
Db 446 ---gvsqignlqtrssqr-----stlyqrnyalnttalcyeapripdyrvq-ecmytr 496
QY 519 SGPALPEGLTASPSIDSIQKDPREFGWRDPELVIOMLQHPSPVOSNAAYLQHL 578
Db 497 lqhvapadgdtctspidsqkprfawrdpelpevlmleqfsgvqanaaayqhlc 556
QY 579 FGDNKIKAEIRROGGIQLVDLDHNMTEVHSAGALRNLYGKANDDKIALKKGCGI 638
Db 557 fgnkvkmevcrilgklhlydlidhvlvqknaqcalnrlvgkstedenklamknv91 616
QY 639 PALVRLLRKTTDLIRLVNGVLMNISCDALKMPIODALVLTNAVLIIPHSQMNPL 698
Db 617 pallrlrlksidaevrelvgvltwnlsscdavkmllirdalstlntvlpnsgwnnsf 676
QY 699 QDDRKIOLHSSQYLRLNATGCLRNVSSAGEARRMECDGLTDLALYVIOALGSSEIDS 758
Db 677 dddhkkikfsgslvrlntgtclnlttsageearqgmsceglvsllyvlhltcnstys 736
QY 759 KTVENCYILRLNLSYRLAETSGQHGKGTDELDCLGAEANGDAASSGCGWKKKKKK 817
Db 737 ktvencvclrlnlsyrllelevpqarllglneldllygkespskdeps-cwqkkkkkk 795
QY 818 --SODQDGVGPIPDCAEPKGIOMLWHPSTIVPYLLTISEGNPTLDECAALNLMA 875
Db 796 tpedqdgvgpripjlskspkgyemlwhpsvvpkyltllaessnpattlesagislqls 855
QY 876 GSNKWSYVIRAAVKEKGLPILVELLRINDRVCAVATALRMALDVRNKEILGRYAM 935
Db 856 snkfaayirgypkrtkglpirlvllmndrvvsgatalrmaldvrnkellygyam 915
QY 936 DLVHRLFGNNSNTASKAMSDOTVAVCTTLEVTIKNNENAKALRDAGIKELVIGISK 995
Db 916 dlvrllpysgnps-----vlsdetmaicallhevsckmenekaladsqyklvnlk 970
QY 996 SKGSHSKYKVAQOYLNSMWQYRDRLSLYKKRGDSQYHFVASSSTIEDRORPVSRR 1055
Db 971 ggrtsrlkvkkaaqvltlwgrydlrsllykdkgwnqnhlfpvstlclerdksh- 1026

```

```

QY 1056 TPSISPV--RVSPNNRSASAPSPREXISLKERKTDYECTGSNATYHAKGERTSRKDM 1113
Db 1027 -pslttngqmspligsvstsspalllgirdrpsaydrtpqpmqyynsgdathk--91 1083
QY 1114 TAQNTGISTLYRNSYGAAPAD-----IKHNOVSQNPVQDESRKDYEQFOKSTFNAYE 1169
Db 1084 ypsdksplylssysppareqnrllhqllyys--qddsnnknfdayrllyqspnshyd 1141
QY 1170 SFPEQVYHHRPPASEYVMHLGLKSTGNVYDFVGAARPYSELANTYSHYAPSDSW 1225
Db 1142 pyfdtrhlf-pastdytqylkstnlyvdfyctkrp-----syraeqypspdsaw 1192

```

RESULT 9  
AAB36464  
ID AAB36464 standard; Protein; 837 AA.

AC AAB36464;  
DT 01-MAR-2001 (first entry)  
XX  
DE Human plakophilin pp2a (PKP2a) protein SEQ ID NO:4.  
XX  
KW Plakophilin-3; PKP3; catenin-like protein; characterisation; diagnosis;  
KW desmosome; epithelial cell; skin disease; dermatological; gene therapy;  
KW vaccine.  
OS Homo sapiens.  
BN WO200066619-A2.  
XX  
PD 09-NOV-2000.  
XX  
PF 28-APR-2000; 2000WO-EP04389.  
XX  
PR 30-APR-1999; 99EP-0870093.  
XX  
PI (VIAA-) VIAAMS INTERUNIVERSITAIR INST BIOTECHNOG.  
XX  
PI Van Roy F, Bonne S;  
XX  
DR WPI: 2000-687529/67.  
XX  
PT Nucleic acids encoding Plakophilin-3 polypeptides, useful for treating  
PT skin diseases and disorders of epithelial tissue associated with  
PT inappropriate Plakophilin-3 expression and activity -  
XX  
PS Example 1; Fig 4; 132pp: English.

The present invention describes an isolated or recombinant nucleic acid molecule (I) encoding a plakophilin-3 (PKP3), from humans, mice and Xenopus laevis. (I) has dermatological activity, and can be used in gene therapy and for vaccines. (I) and the protein it encodes may be used in the prevention, treatment and diagnosis of diseases associated with inappropriate PKP3 expression, such as skin diseases and disorders affecting epithelial tissue. For example, (I) (and vectors containing (I)) and the PKP3 polypeptide may be used to treat disorders associated with decreased PKP3 expression by rectifying mutations or deletions in a patient's genome that affect the activity of PKP3 by expressing inactive proteins or to supplement the patients own production of PKP3 polypeptides. Additionally (I) may be used to produce PKP3, according to standard recombinant DNA methodology, by inserting the nucleic acids into a host cell and culturing the cell to express the protein. (I) and complementary sequences may also be used as DNA probes in diagnostic assays to detect and quantitate the presence of similar nucleic acid sequences in samples, and hence which patients may be in need of restorative therapy. The PKP3 polypeptides may be used as antigens in the production of antibodies against PKP3 and in assays to identify modulators (agonists and antagonists) of PKP3 expression and activity. The anti-PKP3 antibodies and PKP3 antagonists may also be used to down regulate PKP3 expression and activity. The anti-PKP3 antibodies may also be used as diagnostic agents for detecting the presence of PKP3

CC polypeptides in samples (e.g. by enzyme linked immunosorbant assay  
 CC (ELISA)). PKP3 is a catenin-like protein, which is present in desmosomes  
 CC and nuclei of epithelial cells. The present sequence represents human  
 CC PKP2a (pp2a), which is used in the exemplification of the present  
 CC invention.

XX Sequence 837 AA;

Query Match 13.1%; Score 833; DB 21; Length 837;  
 Best Local Similarity 29.8%; Pred. No. 6.1e-51;  
 Matches 259; Conservative 134; Mismatches 323; Indels 152; Gaps 29;

256 SSTLPAPRGSGSPLA--APQGSPTK-----IQRGSAPEG-----ATYAPR- 296  
 27 ssslalpeaklkagssgrgqvtksrldeqvgqtlarqrsyvnqnlhtssvpey 86  
 297 -----GSSPKQSPRLAKSYSTSSPINIVSSAGLSPINVTSPPTVQSTISS 343  
 87 vynlhlvendfvggrspypklydmlkagttat---ygrwgrtqtagysqskveerslr 142  
 344 SPIHLSSTIGTVAFLSPTK-RLVHASROYSK-----HSQLLYATLQLQR-----GS 390  
 143 hpirrlaispds-----speatnhsdyqysqrsqaghtlhqgesrraalvppryarse 198  
 391 LAAGSRASYSQHGHLGPELRALQSPHHIDPIYEDRVYOKP-----PMRSLQSOGD-- 443  
 199 lvyrsagrtarqrhfdtyhryq--hgsyavdtdfslpanalltprpgrtstmgnl 256  
 444 -----PLPRAHGTITRTSTAPSPGVDSVPLORTGSGQHPQNAATAFTQ 487  
 257 ekenyllagtlvgvgrvlpvlqpyt-qnrarseswbqsfhstftrlaeagsvavdsgr 315  
 488 RASYAAGPASNAYADPYRQIQCPSVESPSKSPALPPEGLARSPSIDSIQKDPREFGW 547  
 316 rahltvgaa-----agsgnllteretfcds-----qlgn 346  
 548 RDPE--LPEVQMQ--HQPPVSONAAAYIQLCFGDNKIKAEIRROGIGTOLLVDLDH 603  
 347 admcmtlerawmleahmpsrtsaattfqbccfksaarkrvnqlklqlklv 406  
 604 RMEVHSACGALRNLYGKANDNKTALKNCGIPALVLLKRTDLELRELVTGVLWN 653  
 407 qnevdgravogalnlyf--ednknlevelnqvrlllqvlqvtldlekkqllgllwn 464  
 664 LSSCDALKMPLIDALAVLTNAVLIIPHSQWENSPLODDRKTQ--LHSSQVLRNATGCLRN 722  
 465 lssndklkmlitealltletenllifsgv---pegdykranglllditfynvtgclirm 521  
 723 SSAGEARRRRRCDDGTDLALVYQALGSSSEIDSTVNCYCILRNLSYRLAETS-- 780  
 522 ssagadgrkmtredglidsivhyvgtladypddkatencvcclhnlhsygleaelpek 581  
 781 -----QGHMGTDDELGLGCEANGKDAESSGCGWKKKKKKKKSODOMGVPLDCAE 833  
 582 ysqnilyqnrnltcd-----nkn---sigcfgsrtrkx--eqydv- pmpceks 625  
 834 PPGIGIOMLHPISVYKRYLLLSLSCSNPTLEGAAGALQNLAAAGSMKWSYI--RAAVRKEK 892  
 626 nprgyvelwlnslvlymlylslaksrvnltqeaalgnaltagsmpmpsvagtlvqkes 685  
 893 GLPIVELLALINDRVCANATLRNMALDVRKKEKELGKAMLDVYRLPGCNSNNTWS 952  
 686 glqhtckmhlvgpsvkktaislrlslrlslqnelaketjpdlyslp-----dtyps 740  
 953 KASDDTVTAVCCCTLHEVITKKNENAKALRDAGIEKIGVSKSGDKH--SPKVYVAAQ 1011  
 741 tdlliettasactlmlnlnqnsygnardllntgylkmai--sagdayasmkaskasav 798  
 1012 VLNSMWOYRDLRSLYKKGDSQYHVAS 1039  
 799 llyslwhtelhnaykkaqfkktdfyns 826

RESULT 10  
 AAB36467  
 ID AAB36467 standard; Protein; 821 AA.  
 XX  
 AC AAB36467;  
 XX  
 DT 01-MAR-2001 (first entry)  
 XX  
 DE Xenopus laevis plakophilin-3 protein SEQ ID NO:9.  
 XX  
 KW Plakophilin-3; PKP3; catenin-like protein; characterisation; diagnosis;  
 KW desmosome; epithelial cell; skin disease; dermatological; gene therapy;  
 KW vaccine.  
 OS Xenopus laevis.  
 XX  
 PN WO20066619-A2.  
 XX  
 PD 09-NOV-2000.  
 XX  
 PF 28-APR-2000; 2000WO-EP04389.  
 XX  
 PR 30-APR-1999; 99EP-0870093.  
 XX  
 PA (VLA-) VLAAMS INTERUNIVERSITAIR INST BIOTECHNOG.  
 PI Van Roy F, Bonne S;  
 DR WPI; 2000-687529/67.  
 DR N-PSDB; AAC64794.  
 XX  
 PT Nucleic acids encoding Plakophilin-3 polypeptides, useful for treating  
 PT skin diseases and disorders of epithelial tissue associated with  
 PT inappropriate Plakophilin-3 expression and activity.  
 XX  
 PS Disclosure; Fig 13; 132pp; English.

The present invention describes an isolated or recombinant nucleic acid molecule (I) encoding a plakophilin-3 (PKP3), from humans, mice and xenopus laevis. (I) has dermatological activity, and can be used in gene therapy and for vaccines. (I) and the protein it encodes may be used in the prevention, treatment and diagnosis of diseases associated with inappropriate PKP3 expression, such as skin diseases and disorders affecting epithelial tissue. For example, (I) (and vectors containing (I)) and the PKP3 polypeptide may be used to treat disorders associated with decreased PKP3 expression by rectifying mutations or deletions in a patient's genome that affect the activity of PKP3 by expressing inactive proteins or to supplement the patients own production of PKP3 polypeptides. Additionally, (I) may be used to produce PKP3, according to standard recombinant DNA methodology, by inserting the nucleic acids into a host cell and culturing the cell to express the protein. (I) and complementary sequences may also be used as DNA probes in diagnostic assays to detect and quantitate the presence of similar nucleic acid sequences in samples, and hence which patients may be in need of restorative therapy. The PKP3 polypeptides may be used as antigens in the production of antibodies against PKP3 and in assays to identify modulators (agonists and antagonists) of PKP3 expression and activity. The anti-PKP3 antibodies and PKP3 antagonists may also be used to down regulate PKP3 expression and activity. The anti-PKP3 antibodies may also be used as diagnostic agents for detecting the presence of PKP3 polypeptides in samples (e.g. by enzyme linked immunosorbant assay (ELISA)). PKP3 is a catenin-like protein, which is present in desmosomes and nuclei of epithelial cells. The present sequence represents Xenopus laevis PKP3, as given in the present invention.

XX Sequence 821 AA;

Query Match 12.2%; Score 778.5; DB 21; Length 821;  
 Best Local Similarity 30.0%; Pred. No. 4.8e-47;  
 Matches 240; Conservative 131; Mismatches 315; Indels 113; Gaps 26;

```

QY 329 INVTSPPVQST--ISSPFIHQLSTICTYA-TLSPTRLVNASQSKSHQELATATL 385
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 48 vrakmmqktqtprrvsanysais--ygdltmptqtqysasasysrqlssepki 103
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 386 QR-----PGSLAAGSRASYSOHG-HIGPELR-----ALQSEPHIDPIY- 424
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 104 stgrtsaysvvggggggmsrsrsavdlgqrlsvaagqrgyscstrpnstfhekyp 163
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 425 EDVRVYOKPRMRLSOSQDPL-----PPAHTGYRTSTAPSPGVSVLPQRTSQHQPON 480
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 164 arrdydmslrlrlgdeeryngpaaagsgfygrqvstssg--pvlqrlslnisqpn 220
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 481 AAAATF-QRASYA-----AGPA-----SNVA-----DPY-----ROLQYCP 510
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 221 gggtswvaeaevagrttlrtpamrtlqrfqstnarirlsapygtlntgsgqgmqgmq 280
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 511 SVESYKSKGRLPPE--GTLARSSIDSIOKDREF----- 545
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 281 qmgqsssmngsyvtnmngmhrstapsvrslaesghvgeigldmfdqnkslmsqsfsts 340
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 546 GMRDELPEVIOMLQHOFPSVQSNAAAYLQHLCEFDNKKAEIRROGGIQLVLDLDRM 605
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 341 gtdmdmnpavkhlmadpnlqvaaylqhrctyndseakqarsvgaipkvlfnchn 400
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 606 TEVHRACGALNLYYKANDNKIALKNCGIPALVRLRKTDTLEIRELVYGVLMNLS 665
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 401 qevgrhatgamnllyd--npenkmalveengiyelltaleepdd--elrknytgllwnls 457
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 666 SCDAKMPRIOMALVITNAVYIIPHSQWNSPLQDDRKIQHSSQVLNATGCLNNVSSA 725
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 458 ssdnkharlartltlpltkvlspisgtagasvlgq--nvaseselfynstgflrnissa 514
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 726 GEARRRREDCGLDPAALYVYIOSALSSSEIDSKTYENCVCILRLMRYLAETSQGOMH 785
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 515 spetrgmrecpgllslsvyslsmalqsgskedevanavcrlmlsryllyge-----m 568
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 786 GTDELDGLLCEANGKD--AESGCGWKKKKK-KRSQDQWGVPLPDCAEPKCIQMLW 842
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 569 ppsalrleggg--grstvgdvtgctfpgsrklkeqggadlatfetsrpkqmelllw 627
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 843 HPSIVAPYLTILSECS--NDITLEGAGALQNLAAQSMKW--SYIRAAVKEKGLIIVEL 900
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 628 hpglvallylrllyqcelnkytteeaagalaqrltagrtvaasvlsqvaldgerllpvaldr 687
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 901 LRIDNRVVCATATARNALDVNRNKLIGKYAMRDVLRLEPGNNSNNTASKAMSDDTV 960
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 688 lrtadnnglrlstglrlrnstrnakndemstklvshlllekpa-----dsgeksppadvl 742
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 961 TAVCTLHEVITKNNENAKALRDAGIEKLVGISKSGDKHSPKYVKAASQVLSNKMQR 1020
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 743 viallmlnlttagplaaardivylnglrylkmkdsdpdgaaraacsilltmwys 802
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 1021 DLRSLTKKDGMSQYHVAS 1039
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 803 klneykakgyrkdflsa 821
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

PN W020006619-A2.
XX 09-NOV-2000.
XX 28-APR-2000; 2000MO-EF04389.
XX 30-APR-1999; 99EP-0870093.
XX (VLA-) VLAAMS INTERUNIVERSITAIR INST BIOTECHNOC.
XX Van Roy F, Bonne S;
XX WPI; 2000-687529/67.
XX N-PSDB; AAC64792.
XX Nucleic acids encoding plakophilin-3 polypeptides, useful for treating
XX skin diseases and disorders of epithelial tissue associated with
XX inappropriate plakophilin-3 expression and activity -
XX Example 1; Fig 2; 132pp; English.
XX The present invention describes an isolated or recombinant nucleic acid
XX molecule (I) encoding a plakophilin-3 (PKP3), from humans, mice and
XX xenopus laevis. (I) has dermatological activity, and can be used in
XX gene therapy and for vaccines. (I) and the protein it encodes may be
XX used in the prevention, treatment and diagnosis of diseases associated
XX with inappropriate PKP3 expression, such as skin diseases and disorders
XX affecting epithelial tissue. For example, (I) (and vectors containing
XX (I)) and the PKP3 polypeptide may be used to treat disorders associated
XX with decreased PKP3 expression by rectifying mutations or deletions in
XX a patient's genome that affect the activity of PKP3 by expressing
XX inactive proteins or to supplement the patient's own production of PKP3
XX polypeptides. Additionally, (I) may be used to produce PKP3, according
XX to standard recombinant DNA methodology, by inserting the nucleic acids
XX into a host cell and culturing the cell to express the protein. (I) and
XX complementary sequences may also be used as DNA probes in diagnostic
XX assays to detect and quantitate the presence of similar nucleic acid
XX sequences in samples, and hence which patients may be in need of
XX restorative therapy. The PKP3 polypeptides may be used as antigens in
XX the production of antibodies against PKP3 and in assays to identify
XX modulators (agonists and antagonists) of PKP3 expression and activity.
XX The anti-PKP3 antibodies and PKP3 antagonists may also be used to down
XX regulate PKP3 expression and activity. The anti-PKP3 antibodies may also
XX be used as diagnostic agents for detecting the presence of PKP3
XX polypeptides in samples (e.g. by enzyme linked immunosorbent assay
XX (ELISA)). PKP3 is a catenin-like protein, which is present in desmosomes
XX and nuclei of epithelial cells. The present sequence represents human
XX PKP3, as given in the present invention.
XX Sequence 797 AA:
S0

```

Query Match 12.1%; Score 770.5; DB 21; Length 797;  
Best Local Similarity 29.3%; Pred. No. 1,7e-46;  
Matches 244; Conservative 126; Mismatches 305; Indels 157; Gaps 27;

```

QY 262 PPRGSPILADQGSPTKLRG-----GSAPEGATTAAPRGSSPKOSPSRLASYS 312
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 61 prhngaapepeaeetargtgrgyhtlqagfssrsglsqdklsgtfrplakpayspasws 120
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 313 TSSPINIVVSSAGLSPINVTSPPTVQSTISSPFIHQLSTIGT--YATLSPTKRLVASE 370
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 121 sravdlscsr-----tlss-----ahnggsatgaagyggaqplcpm--ptr 160
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 371 QYKSHQELVATATLQRPGLAAGSRASYSOHGHLPELRALQSPBNHIDPIYEDRYQ 430
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 161 pvsftl-----ergv-----vgsradytl-----slrstlrgpgyddr----- 195
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 431 KPPMRLSOSQGDPLPPAHTGYRTSTAPSPGVDSVPLQRTGSOHQPONAAATFORAS 490
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 196 -----slvseq-----lepatslyra-----fayetgassssraq 227
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 491 YAAGPASNVADPYRLQYCPV-----ESPYSKS-----GPALPPEGLARSPIDS 537
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Db 228 glDpeatevpsrtlr-apaavrclgrfqshtsrvgavpavlepp---varapvrs 283
OY 538 IQ-----KDPREF-----GMRDPELPEYIOMLQHOFPSVGSNAAYL 574
Db 284 lslsladgqlpdvghfnsyshrllqrlssgldldlpsavkylmsdplqlvlgayl 343
OY 575 QHLFCGDNKIKAEIRROGGIOLVLDLDRHMTVEHRSACGALRLVYGRKANDPKIKLN 634
Db 344 qhkcydaaakqgslqavprlvklfnhangeqrnatgamrllld--nadrkklive 401
OY 635 CGGIPALVRLRLKRTTDLFIRELVTGVLNLSGCDALKPITQDALVLAIVNAVTPHSWE 694
Db 402 enqifellrtleqdd-elrknavgllwnlsschldkrlardleqldtlvlspsgag 460
OY 695 NSPIODDRKILHSS--OVLNRTGCLRNVSSAGEARRRMRREDGTLTDLALVYIOSALG 752
Db 461 gppl-----lqnasaeelfynaqlftrlnssasqatrqkmrtechglvdaivtslnald 515
OY 753 SSEIDSKTVENCVCILRLNLSYRL-----AAETSGOHMGTDLDLGLCEANGKDAES 805
Db 516 agkedcsvenavcvinlsyrllydempssalgrlegr--grldlagappegv----- 566
OY 806 SGCGKKKKKKKSSODQDVGVRPDLCAEPKGIOMLHNPSTVAKYLLLESCS-NPDTLE 864
Db 567 vgcftgqrrrlreiplaadaltfaevskdpxglewlsptqlvlylrlgrcelnhttle 626
OY 865 GAAGALONLAGSWKWS-VYTRAAVREKGPILVELLRIDNDRVCAVATLRNMLAD 923
Db 627 aaagalnigtadgrwagvlsrlalegerlrlpldrctadqhqlslglllnsrna 686
OY 924 RNKELIGKYAMRDLYVHRLPGNNNSNTASKAMSDDTVAVCCTLHEVITRNMENAKALRD 983
Db 687 rnkdemstkvshlklelpq-----svgekspaeavlwnllavlnnlvasplaaadlly 741
OY 984 AGGIEKLVGISKSGDKHSPVYKASOVLSNMVYRDLSIKYKDDQMSQYHYV 1037
Db 742 fdgrlrkllfllkkrrdspdsekssraassllanlwqynklndlrakxyrkedtl 795

RESULT 12
AAB36466
ID AAB36466 standard; Protein; 797 AA.
AC AAB36466;
AD
DE 01-MAR-2001 (first entry)
XX
XX Mouse plakophilin-3 protein SEQ ID NO:7.
XX
XX Plakophilin-3; PKP3; catenin-like protein; characterisation; diagnosis;
XX KW desmosome; epithelial cell; skin disease; dermatological; gene therapy;
XX vacuome.
XX OS Mus musculus.
XX
XX WO200006619-A2.
XX
XX 09-NOV-2000.
XX
XX 28-APR-2000; 2000MO-EP04389.
XX
XX 30-APR-1999; 99EP-0870093.
XX
XX (VLAAs) VLAAms INTERUNIVERSITAIR INST BIOTECHNOG.
XX
XX Van Roy F, Bonne S;
XX
XX WPI; 2000-687529/67.
XX DR N-PSDB; AAC64793.
XX
XX Nucleic acids encoding Plakophilin-3 polypeptides, useful for treating
XX PT skin diseases and disorders of epithelial tissue associated with

```

```

PT Inappropriate Plakophilin-3 expression and activity -
XX Disclosure: Fig 11; 132pp; English.
XX
XX The present invention describes an isolated or recombinant nucleic acid
XX molecule (I) encoding a Plakophilin-3 (PKP3), from humans, mice and
XX Xenopus laevis. (I) has dermatological activity, and can be used in
XX gene therapy and for vaccines. (I) and the protein it encodes may be
XX used in the prevention, treatment and diagnosis of diseases associated
XX with inappropriate PKP3 expression, such as skin diseases and disorders
XX affecting epithelial tissue. For example, (I) (and vectors containing
XX (I)) and the PKP3 polypeptide may be used to treat disorders associated
XX with decreased PKP3 expression by rectifying mutations or deletions in
XX a patient's genome that affect the activity of PKP3 by expressing
XX inactive proteins or to supplement the patients own production of PKP3
XX polypeptides. Additionally, (I) may be used to produce PKP3, according
XX to standard recombinant DNA methodology, by inserting the nucleic acids
XX into a host cell and culturing the cell to express the protein. (I) and
XX complementary sequences may also be used as DNA probes in diagnostic
XX assays to detect and quantitate the presence of similar nucleic acid
XX sequences in samples, and hence which patients may be in need of
XX restorative therapy. The PKP3 polypeptides may be used as antigens in
XX the production of antibodies against PKP3 and in assays to identify
XX modulators (agonists and antagonists) of PKP3 expression and activity.
XX The anti-PKP3 antibodies and PKP3 antagonists may also be used to down
XX regulate PKP3 expression and activity. The anti-PKP3 antibodies may also
XX be used as diagnostic agents for detecting the presence of PKP3
XX polypeptides in samples (e.g. by enzyme linked immunosorbant assay
XX (ELISA)). PKP3 is a catenin-like protein, which is present in desmosomes
XX and nuclei of epithelial cells. The present sequence represents mouse
XX PKP3, as given in the present invention.
XX
XX Sequence 797 AA;
SQ
Query Match 12.0%; Score 766; DB 21; Length 797;
Best Local Similarity 27.5%; Pred. No. 3,66-46;
Matches 266; Conservative 125; Mismatches 380; Indels 196; Gaps 26;
OY 85 LSGETGSMSSMSAEQFOWQSGDKDIDELTTGLVNSCIRLSQESGILDPDYST 144
Db 11 lqpetgvcsalpsldldrrgaegpe--adrlaarvqeyarllly-----lqgsrlm 64
OY 145 GERPSLLSQSALQJNSKEPGRFYPASVHNSQTLALGETPSQLPARKQARATQGSFSQ 204
Db 65 gs--aelgsaesargmprry-----ghlmqt-----gfsersgmsgdK 102
OY 205 GTTSRAGHLAGEPAPPPPPPPREFPAPSLGSARELPAPRAAAALYSSSTLPAPPR 264
Db 103 tsllr-----plakpayspawssrsavdltsrllssahnggsafigavvy 148
OY 265 GSPSLAPQGGSPTRLQGRGSAPEGATYAAPRGSPKPSRLAKSYSTSPINIVYSSA 324
Db 149 ggtcplpmptrpvsfhergaaasradydtslpslltgggldlrys-----vsse 200
OY 325 GISPRLVTSPTVOSTISSPFIHQLSTTIGYATLSPTKRLVHNASQYSKHSDLYATAT 384
Db 201 qlpe-----aaastlyraya-----ykrqass----- 221
OY 385 LQPGSLAAGSRASYSQGHGLGPELRALQSPENHIDPIYEDRYQKPPAKSLISQSGDP 444
Db 222 ----gssrag-----glDpeategr--psrlirapantllqrfqsh 258
OY 445 LPPAHTGTYR---TSTAPSPGVDSVPLQRTGSGOHGQNMAAAATFGQASVYAAGPASYAD 501
Db 259 rsrgtgvsvsgaggleparavsvrslsladsgn----- 293
OY 502 PYROLQYCPVSESPYSKSGPALPEGTLARSPSIDSIQKDPREFGMRDPELPEYIOMLQH 561
Db 294 -----lpdvrgldsyg-----hrlqrls-----gfddldpsavkylma 330
OY 562 QPSPVGSNAAYLOHLCEGDNKIKAEIRROGGIOLVLDLDRHMTVEHRSACGALRLVY 621

```







DR MPI: 1999-561617/47.  
DR N-PSDB: AA223430.

XX New proapoptotic dependence peptides, used to develop products for  
PT treating, e.g. Alzheimer's disease -  
XX  
PS  
XX

Disclosure: Page 154-157; 1999p; English.

CC This invention describes novel pure proapoptotic dependence peptides  
CC which comprise a sequence of an active dependence domain selected from  
CC dependence polypeptides consisting of p75NTR, androgen receptor, DC,  
CC huntingtin polypeptide, Machado-Joseph disease gene product, SCAL, SCAL2,  
CC SC6 and atrophin-1 polypeptide. The proapoptotic peptides are capable  
CC of inducing cell death and can be used to develop products to mediate or  
CC inhibit apoptosis. The methods can be used for reducing the severity of  
CC Huntington's disease, Alzheimer's disease, mediated pathological conditions e.g.  
CC Spino cerebellar ataxias, dentatorubropallidoluyian atrophy,  
CC Machado-Joseph disease, stroke or head trauma. They can also be used for  
CC reducing the severity of a pathological condition mediated by upregulated  
CC cell proliferation or cell survival e.g. neoplastic, malignant,  
CC autoimmune or fibrotic conditions. This sequence represents the human  
CC atrophin 1 polypeptide described in the method of the invention.  
CC  
XX  
SQ

Sequence 1185 AA;

Query Match 4.0%; Score 256.5; DB 20; Length 1185;  
Best Local Similarity 20.8%; Pred. No. 2e-09;  
Matches 180; Conservative 87; Mismatches 268; Indels 329; Gaps 40;

```

QY 32 SP-GLNTSNGSGSETETSAIILASVKEQLFRLTRELAEKQIVASQLERCKUGSETG 90
DB 34 spgsvstssadg-kaeksvrtakkarveastpkvnggrse-elseseetnapkktk 91
QY 91 SMSMSAAEFOFQWQODGQKDIETLGTGLVDSICRLQESGILPDQYSTGR--- 147
DB 92 tegelprpgspaldsidg-----rslndgssprdidqdnrtss 132
QY 148 PSLISQSALQINS-----KPEGSFQ-----YPA 170
DB 133 pslypsvsvendsssgslsgparpyhnpplfpssppdsctprqpeasfephsvtrpt 192
QY 171 SYHSNQ---TLALGETTPS-----OLPARGTQARATGQSFQ---GTSRAGHLAGEEP 218
DB 193 gylhamemprtsrnfqppgapphpqlypgygtvgvlsppmmpkkggaaavsgpnpqkx 252
QY 219 APP-----PPPPPEFAPSLGSAFHLPDAPPA-----AAAL 252
DB 253 hppptprlsvsssgaagappctkptctvggg--nlpsappanfpvtrpnlpppalrpl 310
QY 253 YSSSTLP---APPRG---SPLAFO---GGSPTKLQRGSAPEGATYAAPRGSSPKOSP 304
DB 311 masasaprpqlgaqplpghlpspyamqgmglppgpekypcllappshlppasssaparp 370
QY 305 SRLAKSYTSSPI---NIVSSAGLSPIRYTS-----PPTVQSTISSPIHQLSS 351
DB 371 mrfpysssssssssssssssssssaaprfpaqalpsyphtfpptslsvsnqpkylqr 430
QY 352 TIGTYATLS-----PTKRLVHAS-----EQ 371
DB 431 slpsqgavsqgpppppygrllansnabpfpstcgastahppvstlhhhhgqgqgq 490
QY 372 YSKHSGELATATLORPGS---LAAGSRASVSSOHG---LGPRLRLQ---SPEH--- 418
DB 491 qgqgqgqhngsgppppgafrhplieg---sshhahpamspslglrpyppgahlp 545
QY 419 --HIDPIYEDRVYQKPEKMSLSQ---SOGD-----PLPAHT--- 450
DB 546 pphagvyysgaqnpqpyvssssssstsgsypscshpspsgpgagaypfpvptvltss 605
QY 451 -----GTYRTSTAPS--SPGVDSVPLQRTG 473

```

```

DB 606 satlstviatvasspagyktaspppppykrapspgayktatpgykpg--spsfirtg 663
QY 474 SOHGPQN---AAATFQORASYAAGPASNADPYROLQYCPVESPYSKGPA-----LP 524
DB 664 tppgyrgtspapsgtffkpsptvgp-----gplppagsglpslpp 705
QY 525 PEGTLARSP--SIDSIKDPREFGWRDEL-----PEYIOMLQHPSPVQSNNA 571
DB 706 ppaapasgplsatqlkxepae-eyetpsvppparspspkvtvgv-----psllasgsa 759
QY 572 AYLOHLCFGDKIKAEIRROGGIOLIVLDLHMTVEHRSACGALRNLYYGKANDDNKTA 631
DB 760 rfukhltrgfn-----sca--tsdlyfvplplegska 788
QY 632 LKNGGIPALVRLRKTTDLRNE 655
DB 789 kkrad---lvekvrreaeqrare 808

```

Search completed: July 19, 2001, 16:03:18  
Job time: 188 sec

Fri Jul 20 10:09:39 2001

us-09-501-171a-4.rag

---

Page 16

1	256.5	4.0	1185	4	US-09-041-886-23	Sequence 23, Appl
2	194.5	3.1	2842	1	US-07-741-940-7	Sequence 7, Appl1
3	194.5	3.1	2842	1	US-08-289-548A-7	Sequence 7, Appl1
4	194.5	3.1	2842	1	US-08-452-654-7	Sequence 7, Appl1
5	194	3.0	2843	1	US-08-452-655B-2	Sequence 2, Appl1-2
6	194	3.0	2843	1	US-08-452-655B-7	Sequence 7, Appl1
7	194	3.0	2843	4	US-08-450-582-2	Sequence 7, Appl1
8	194	3.0	2843	4	US-08-450-582-7	Sequence 7, Appl1
9	194	3.0	2973	2	US-08-821-355A-7	Sequence 7, Appl1
10	194	3.0	2973	2	US-09-003-687A-7	Sequence 7, Appl1
11	194	3.0	2973	4	US-09-136-605-7	Sequence 7, Appl1
12	194	3.0	2843	1	US-07-741-940-2	Sequence 2, Appl1-2
13	192	3.0	2843	1	US-08-289-548A-2	Sequence 2, Appl1-2
14	192	3.0	2843	1	US-08-432-654-2	Sequence 2, Appl1
15	192	3.0	2843	2	US-08-370-235A-2	Sequence 2, Appl1
16	191	3.0	1068	1	US-08-396-479B-12	Sequence 12, Appl1-2
17	191	3.0	1068	1	US-08-818-823-12	Sequence 12, Appl1
18	183	2.9	1004	4	US-08-916-352-2	Sequence 2, Appl1
19	180	2.8	1481	2	US-08-616-844-40	Sequence 40, Appl
20	180	2.8	1481	2	US-08-559-654-40	Sequence 40, Appl
21	180	2.8	1481	3	US-08-944-868A-40	Sequence 40, Appl
22	180	2.8	1481	3	US-08-944-423A-40	Sequence 40, Appl
23	180	2.8	1481	4	US-08-944-496-40	Sequence 40, Appl
24	175	2.7	1075	5	PCT-US94-07297-41	Sequence 41, Appl
25	174.5	2.7	1005	2	US-08-935-450-2	Sequence 2, Appl
26	173	2.7	984	4	US-08-764-870-15	Sequence 15, Appl
27	171	2.7	1142	2	US-08-953-118-7	Sequence 7, Appl1

28	171	2.7	1142	3	US-08-845-528C-7	Sequence 7, Appl1
29	169	2.7	509	4	US-08-860-635A-11	Sequence 21, Appl1
30	168.5	2.6	708	1	US-08-396-479B-8	Sequence 8, Appl1
31	168.5	2.6	708	1	US-08-818-823-8	Sequence 8, Appl1
32	168.5	2.6	739	1	US-08-846-479B-10	Sequence 10, Appl1
33	168.5	2.6	739	1	US-08-819-862-10	Sequence 10, Appl1
34	165.5	2.6	1003	1	US-08-571-758-4	Sequence 4, Appl1
35	165.5	2.6	1003	1	US-08-909-988A-4	Sequence 4, Appl1
36	165.5	2.6	1003	1	US-08-909-983-4	Sequence 4, Appl1
37	159.5	2.5	534	2	US-08-878-563A-1	Sequence 1, Appl1
38	158.5	2.5	1321	2	US-08-817-310A-04	Sequence 64, Appl1
39	157.5	2.5	1317	3	US-09-083-521-7	Sequence 7, Appl1
40	157	2.5	1312	4	US-09-041-881-19	Sequence 19, Appl1
41	153.5	2.4	2441	1	US-08-194-468-2	Sequence 2, Appl1
42	155.5	2.4	2441	1	US-08-661-739-2	Sequence 2, Appl1
43	153.5	2.4	737	1	US-08-188-583-16	Sequence 16, Appl1
44	153.5	2.4	737	1	US-08-646-715-16	Sequence 16, Appl1
45	152	2.4	507	4	US-08-860-635A-19	Sequence 19, Appl1

## ALIGNMENTS

RESULT 1  
US-09-041-886-23

Patent No. 6235872

APPLICANT: Bredesen, Dale E.

TITLE OF INVENTION: Proapoptotic Peptides, Dependence

NUMBER OF SEQUENCES: 72

ADDRESSEE: Campbell & Flores LLP

CITY: San Diego

COUNTRY: United States

COMPUTER READ

COMPUTER: IBM PC compatible

SOFTWARE: PatentIn Release #1.0, Version #1.25

APPLICATION NUMBER: US/09/041,886

CLASSIFICATION:

NAME: Campbell, Cathryn A

REFERENCE/DOCKET NUMBER: P-LJ 2626

TELEPHONE: (619) 535-9001

INFORMATION FOR SEQ ID NO:

LENGTH: 1185 amino acids

TOPOLOGY: linear

US-09-041-886-23

Query Match	4.08; Score 256.5; DB 4; Length 1185
-------------	--------------------------------------

Matches 180; Conservative 87; Mismatches 268; Indels 329; Gaps 40;

QY 32 SP-GLNTSNGDSETETTSAILASVKEQELQFERLTRELEAERQIVASQLERCKLGSETG 90

Db 34 SPGCVSTSSDG-KAEKSRQAKKARVEASTPKVNKQGRSE-EISESESEETNAPKTK 91

```

QY 91 SMSSSSAEFOFOWOSODGKDIEDLTGTLGLVDSICIRLSQESGILDPODYSTER--- 147
: : : : :
Db 92 TEQELPRFQSPDSDLDLDG-----RSINDGSDPDIDDNSTIS 132
: : : : :
QY 148 PSLLSQSLQJLNS-----KREGSFQ-----YPA 170
: : : : :
Db 133 PSIVSPGSEVENDSDSSGSLSGPARPYHPPLFPSPPOPPDSTPRQPEASEFEPHSPVPT 192
: : : : :
QY 171 SYHSNO---TALGETTPS-----OLPARGTQARATGQSFQ---GTSNAGLAPREP 218
: : : : :
Db 193 GYHAPMEPTSRMFOAPRGAPRPPRPPOLYRGSTGVLSGPPMGKRGGAASVGGPNSGKO 252
: : : : :
QY 219 APP-----PPPREPAPSLGSFHLDPAPPA-----AAAL 252
: : : : :
Db 253 HPPTPTIVSSGASGAPRPPKPTTPVGGG--NLPSAPPANPHTVPLPPLPPALRPL 310
: : : : :
QY 253 YSSSTLP---APPRG---SPLAPQ---GSPPTKLDGSSAPRGATYAAROGSSPQSP 304
: : : : :
Db 311 NMAASAPGICAGPLFGLHPSPYMGQMGGLPPGKPGTLAPSPHSLPSPASSAPAP 370
: : : : :
QY 305 SRLAKSYSTSPI-----NIYVSAGLSPIRYTS-----PPTVQSTISSPIHQSS 351
: : : : :
Db 371 MRPFYSSSSSSAAASSSSSSSSSAPFRASQALPSTPHSPPTTSLVSNQPKYTOP 430
: : : : :
QY 352 TIGYATLS-----PTKRLVHAS-----EQ 371
: : : : :
Db 431 SLPSQAVWSQGP PPPPPYGRLLANSNAHPFPPTGAGSTANPPVSTHHHHHQQQQQQ 490
: : : : :
QY 372 YSKHSOELYATLQRPSS---LAAGSRASYSQGH---LQELRALQ---SPH-- 418
: : : : :
Db 491 QQQQQQHHGNSGPPPPGAPRPHLEGG---SSHNAHYAMSPSLGSLPPYPPGPAHLR 545
: : : : :
QY 419 --HIDPIYEDRYOKPPMRLSQ-----SOGD-----DLPRAHT--- 450
: : : : :
Db 546 PPHGOVSYSQAGPCPGPPVSSSSSSSTSGCSYPCSHSPSQGQAGPYFFPPVPTTIS 605
: : : : :
QY 451 -----GTYTSTAPS--SPGVDSVPLQRTG 473
: : : : :
Db 606 SATLSTVIATVASSPAGYKTASPPPPYGRAPSPGAYKTATPPGYKPG--SPPSFRTG 663
: : : : :
QY 474 SQHGPN-----AAATFORASYAGPASNADPYRLQYCPVESPYKSGPA-----LP 524
: : : : :
Db 664 TPPTRTGTPRAGCTGPKGSPYGP-----GPPPGPFGSLPLP 705
: : : : :
QY 525 PEGTLANSP--SIDSIKDPREFGWRDEL-----PEVYOMLQHPFSVSNAA 571
: : : : :
Db 706 PPAAPASGPPLSATQIKQEPAE--EYTPESPVPAPRSPSPKVVY-----PSHASQSA 759
: : : : :
QY 572 AYLOHGFQDNKIKAEIRROGGLVLDLDHRTETVHRSCAGCALRLVYKANDNKIA 631
: : : : :
Db 760 RENNHLDRGF-----SCA--RSDLYFVPLEGSKLA 788
: : : : :
QY 632 LKNCGIPALVRLKTTLEIRE 655
: : : : :
Db 789 KKRAD---LVEKYRREAEOARRE 808
: : : : :

```

RESULT 2  
US-07-741-940-7  
; Sequence 7, Application US/0741940  
; Patent No. 5352775

GENERAL INFORMATION:  
; APPLICANT: ALBERTSEN, HANS  
; APPLICANT: ANAND, RAKESH  
; APPLICANT: CARLSON, MARY  
; APPLICANT: GRODEN, JOANNA  
; APPLICANT: HEDGE, PHILIP J.  
; APPLICANT: JOSIYN, GEOFF  
; APPLICANT: KINZLER, KENNETH  
; APPLICANT: MARKHAM, ALEXANDER F.  
; APPLICANT: NAKAMURA, YUSUKE  
; APPLICANT: THLIVERIS, ANDREW  
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC

```

: TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
: NUMBER OF SEQUENCES: 94
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Banner, Birch, McKie & Beckett
: STREET: 1001 G Street, NW
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20001-4598
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/07/741,940
: FILING DATE: 1992/01/09
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Kagan, Sarah A.
: REGISTRATION NUMBER: 32,141
: REFERENCE/DOCKET NUMBER: 1107.035574
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-508-9299
: TELEFAX: 202-508-9299
: INFORMATION FOR SEQ ID NO: 7:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2842 amino acids
: TYPE: AMINO ACID
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: ORIGINAL SOURCE:
: ORGANISM: Homo sapiens
: IMMEDIATE SOURCE:
: CLONE: APC
: US-07-741-940-7

Query Match
Best Local Similarity 3.1%; Score 194.5; DB 1; Length 2842;
Matches 232; Conservative 164; Mismatches 428; Indels 431; Gaps 49;

QY 23 SASERTSLSP-----GLNTSNGDGESE--TTSAILASVKEDE-----LQ 61
: : : : :
Db 102 SVSSRSGECPVPMGSPFRGFGVNGSRSTGYLELEKERSILLADLDKEKEDWYAAQ 161
: : : : :
QY 62 FRLTRLEEA-----ERQIVASQLE-----RCKLGETGSMSSMSAEQFOWOSQ 107
: : : : :
Db 162 LQNLTKRIDSLTFENESLQTDWTRRLQLEYEARQITVAMEEQIGTCQDEKRAQRRIARIQ 221
: : : : :
QY 108 DQGDIEDLTLGLVDSICIRLSQESGILDPODYSTERPSLLSQSLQNSKPEGSFQ 167
: : : : :
Db 222 QLEKDI-----LRITROLQS-----QATEHERSS-----QNKHETGSHD 255
: : : : :
QY 168 YPASHNSQTLALGETTPSQLPARGTQARATGQSFQGTTSRAGHLAGEPAPPPPPR 227
: : : : :
Db 256 ---AERONEGQGVGEIN-----MATSGN--GQGSTTRMDH----- 285
: : : : :
QY 228 EPPAFSLGSFHLDPAPPAALAIYSSSTLPPAPRPGSGPLAPQGGSPYKLDGGSAP 287
: : : : :
Db 286 -----ETASVLSSSSTHSAPRRLTSHLG-----TKVEMVYS-- 316
: : : : :
QY 288 EGATYAAPRGSSPKOSPRLAKSYSTSPINIVVSAGLSPIRYTSPPTVQSTISSPIH 347
: : : : :
Db 317 ---LLSMGLTHDKDMKRTLLAASSSDSCISKRQSGCLPLI----- 356
: : : : :
QY 348 QLSSTIGYATLSPYKRLVHASQYKSHSQELVYATATLQRPGLAAGSRASYSQHG--H 405
: : : : :
Db 357 -----QLHGNNDKS-----VLGNSRGSKEARARS--AAHNIH 391
: : : : :
QY 406 LGPE-----LRALQSEHHIDPIYEDRYVOKPPMRLSOSQGDPLPRAHTGYRT 455
: : : : :

```

```

Db 392 SQPDKRGREIRIVLHLLDQIRAYCETCMEQEAHERGM-----DQDKNMP----- 438
Qy 456 STAPSSPVDSVPLQRTGSOHQHQAATAATFORASYAGPASNYADPYRQLOYCPSEVSP 515
Db 439 -----APVEH-----QICPAV----- 449
Qy 516 YSKSGPALPREGTLAAPSIDSIOKDP-REFGMRODELPEVIOMLQ-----HOE 563
Db 450 -----CVLMKLSFDEEHRHAMELG-----GLOAIALLQVDCEMYGLTNDHS 493
Qy 564 PSYQNAAYLOHLFCGDKIKAEI-RROGGIQLVDLIDHRTVEYHRSAGCALNLVYG 622
Db 494 ITRRRAGALNLTPTGDVANKATLCSMGCBALVLAOLKSESEDIQOYIAYVNLNLS- 552
Qy 623 KANDKIALKNCGGIPALVRLKRTTDEIRELYTVGLNLS-CDALKMPT--IODAL 679
Db 553 RADVNSKKTIREVGSYKALMECALEYKRETLKSVLSALMNLJAHCTEKKADICAVDAL 612
Qy 680 AVLTNAVITPHSGWENSPLODDRKTOLHSSQVLRNATGLRVSS--AGEARRMRREC 736
Db 613 AFLVGLTY-----RSQTNLTAIESGGILRVNSSLIATNEDHROIAREN 658
Qy 737 DGLTDLALYVIOALGSESEIDSKTYENCYCILNLSYRLAETSGQHMGTELDGLCG 796
Db 659 NCLQTLLOHLKSHL-----TYSNAGCTLMNLSAR----- 689
Qy 797 EANGKADESSGCGKKKKKKKQSDQMDGVPLDCAEPPKGIOMLHPSIVKPYLTLSSE 856
Db 690 --NPKQEA-----LMDGAVS-MLKNIJH 711
Qy 857 CSNPDLLEGAAGLQTLA-----GSWKSIVYIRAAVKKKGPILVELL- 901
Db 712 SKHKMTAMGSAALRLMANNRPAYKADANIMSPGSLPFLH--VRKQALEAEALQOH 767
Qy 902 -----RIDNRYVCAVATALRNMALDYANKELLIGKYARDLVHRLPGGNNNTASKAM 955
Db 768 LSEFPNDIN-----LSPKASHRSQNRHKQSLYGDYF-----DINRHDNRSDNF 813
Qy 956 SDDTYAVACCTHEVYTKKMEAKALRDAGIEKLVGISKSG-----D 999
Db 814 NNGMNTVLPSPYNTVTLVPSRSSSGSLDSSREKDSLERERGTGLGNYPATEMPTSS 873
Qy 1000 KISPKVYKASQVLYNSMGOYRDLRLSYK--KDGWQYHFAVASSSTIENDRPRPSSST 1056
Db 874 KRGLQSTTAQAIKAYMEVSAIHTSQEDRSSGTTTELHCY--TDENNALRRSSAHT 929
Qy 1057 PS--ISPVSEPNNSASAPASPREMISLKERKTDYECTGSNATYHAKGEHTSKDAMT 1114
Db 930 HSNVTYFTSENSNRCTCSMPYAKLE--YKRSSNDSLNSVSSSDGYCKRGQ--MKPSIE 983
Qy 1115 AONTGISTLYRNSYCAPADEDIKHNQVSAQVPQEPSRKDYETYPONSTRNYDE 1169
Db 984 SYSEDESKF-CSYGGYPADLAKHKIHSANHMD--NDGELDTPLNVLKYSDE 1033

RESULT 3
US-08-289-548A-7
; Sequence 7, Application US/08289548A
; Patent No. 5648212
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 102

```

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Allegretti, LTD
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/289,548A
FILING DATE: 12-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107,46943
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2842 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: APC
US-08-289-548A-7

Query Match 3.1%; Score 194.5; DB 1; Length 2842;
Best Local Similarity 18.5%; Pred. No. 5,1e-05;
Matches 232; Conservative 164; Mismatches 428; Indels 431; Gaps 49;

Qy 23 SASKTSLSF-----GLNTSNGDSETE-TTSAILASVKEDE-----LQ 61
Db 102 SVSSRGSESPVMPGSPRRGFGVNGSRSTGYLEELKERSLLADLDKKEKEDMYA 161
Qy 62 FERLTRELEA-----EROIVASOLE-----RCKLGSETGSMSSSAEFOGWSQ 107
Db 162 LQNLTKRIDSLTENFSIQDMTRQLFEARQIRVAMERQLGTCQDMKRAQRIARLQ 221
Qy 108 DGOKDIEDLTTGLVLDSCIRSLQESGILDPDYSTERPSSLSQSAIQLNSKPGESFO 167
Db 222 QIEKDI-----LRIRQLQS-----QATEAERSS-----QNHETGSHD 255
Qy 168 YPASYHNOILALGETTPSQLPANGTOARATGOSFSQGTTSRAGHLAAPPAPPPPPPP 227
Db 256 --AERONEGQGVGIN-----MATSGN-GGGSITRMDH----- 285
Qy 228 EPFAPSLGSAFHLPDAPPAAMAAALVYSSSTLPAPPGSGSPLAPOGGSPTKLRGGSAP 287
Db 286 -----ETASVLSSSSTHSAPRRLTSHLG-----TKENYYS-- 316
Qy 288 EGATYAAPRGSSPKQSPSLAKSYSTSSPINIVVSSAGLSLPINVTSPPTVQSTISSPFI 347
Db 317 ---LTLMLGTHDKDDMSRLTLMSSSSQSDSCISWRGSGCLPLLI----- 356
Qy 348 QLSSTIGTYATLSPTRKRLYHASQYSKHSELYATATLQRPGLASGRASYSQHG--H 405
Db 357 -----QLLHGNDKS-----VLGNSRGSKEARARS--AALNIIH 391
Qy 406 LGPE-----LRLAQSPENHIIDPIYEDRYVOKPMPMSLSQSGODPLPPAHTGTYRT 455
Db 392 SQPDKRGREIRIVLHLLDQIRAYCETCMEQEAHERGM-----DQDKNMP----- 438

```

```

QY 456 STAPSPGVDSVPIORRGSGHQPNMAAATPQRASVAAAPASVADPPYRQLOVCPSEVESP 515
DB 439 -----APVEH-----QICPAY---449
QY 516 YSKSGPALPREGTLARSPIDSIOKDP-REFGWRDPPELVYIOMIQ-----HOE 563
DB 450 -----CYLMKLSFDEERHANNELG-----GLOAIETLLQVDCEMVGLTNDHYS 493
QY 564 PSVOSNAAYIOLHCFQGNKIKAEI-KRQSIGIOLLVLDLHRMTFVRSACGALRLVYG 622
DB 494 ILLRYAAMALTNLTFFGVANKATLCMSKGMCAVQALVQLESEBDLQOVLASVLRNLMSW- 552
QY 623 KANDNKITALKNCGIPALVRLKRTDLEIRELVGLVNLSS-CDALMKPI--IQDAL 679
DB 553 RADVNSKTLREVGSKALMECALVEKKESTLVLSATMLMSLHCHENKADICAVGAL 612
QY 680 AVLINAVIIPHSWENSLQODRKIQHSSOVLNANAGCLRNVS---AGEARRRREC 736
DB 613 AFLVETLLY-----RSQNTLALIESGGILRNVSLSLATNEDHROLREN 658
QY 737 DGLFDALYVIOALGSSEIDSKTYENCVCILRNLSYRLAETSOQOHMGTELDGLLG 796
DB 659 NCLQTLLOHLKLSHL-----TIVSNACGTLMLMSAR-----689
QY 797 EANGKDAESSGCGWKKKKKKKSDQMDGVGPLPCAEPPKGIOMLHPSIVKPYLLTLE 856
DB 690 --NPKDQEA-----LMDGAVS-MLKNLIH 711
QY 857 CSNPTELEGAALQNLAA-----GSMKKSIVYIRAVKREKPLTVLLEL- 901
DB 712 SKHKMIANGSAALRNLMANRPARYKKNANIMSPSSSLPSLIH---VRKQALEALDAQH 767
QY 902 -----RIDNRPVCAVATLARNMALDVNRKELIGKYAMRDVLRHLPGGNSNNTASKAM 955
DB 768 LSEFTDNIDN-----LSPKASHSKQHRKQSLVGDYF-----DTNHHDNRDNEF 813
QY 956 SDDVTAVCCCTLHEVITKMNENAKALDAGIEKLVGISKSG-----D 999
DB 814 NTGNNTVLSPLYNTVLTPLSSSSSRGSLDSRSEKDRSLERBRIGLGNYHPATENPGTSS 873
QY 1000 KHSRPNVAAAOVLNSMOYQDIDSLYK---KDKMSQYHFPAVASSSTIERDQORYSSRT 1056
DB 874 KRGLQISTTAAQIAKVMEEVSAIHISQEDRSSGSTTELVHCV---TBERNLRLRSSAAHT 929
QY 1057 PS--TSPRVAPGNNSAAPASPREMISLKERTDYECTGSNATYHGAKGHTSRKDPAMT 1114
DB 930 HSNNTNFTKSEKSNRTCGMPYAKLE---YKRSSNDLSNVSSSDGYGKRQ---MKPSIE 983
QY 1115 AONTGISLTYRNSYGAPAEEDIKHNQVSAQVPQPEPSRKDYETYPQONSTRNDE 1169
DB 984 SYSEDESKF-CSYGOYPADLAHKIHSANHMD---NDGELDTPINYSKYSDE 1033

RESULT 4
US-08-452-654-7
; Sequence 7, Application US/08452654
; Patent No. 5691454
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THILVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie & Beckett

```

```

; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452,654
; FILING DATE: 25-MAY-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/741,940
; FILING DATE: 08-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107,035574
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2842 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: APC
US-08-452-654-7

Query Match 3.1%; Score 194.5; DB 1; Length 2842;
Best local similarity 18.5%; Pred. No. 5,1e-05;
Matches 232; Conservative 164; Mismatches 428; Indels 431; Gaps 49;

QY 23 SASKTSLSLP-----GLTNSNGDSETE-TTSAIIASVKEQ-----LQ 61
DB 102 SVSRSGDESPVPMGSPFRGFGVNGSRESTDYIELEKERSLLDLQKEEKDYYAQ 161
QY 62 FERLTRELEA-----ERQIYASOLE-----RCKLSETSMSMSASAEOPOMOSO 107
DB 162 LQNLTRIDSLTLENSLQTDMTTRQLEYEARQIRYAMEQUGTCODMEKRAQRIARIQ 221
QY 108 DGQKIDEDLTGELVDSQIRSLQESGILDPQDYSTGERPSILSQSALQINSKEGFSQ 167
DB 222 QIEKDI-----LRIRQLQS-----QATEARRS-----QKHEHGSND 255
QY 168 YPASYNSNOTLIGETTPSQLPARGTQARATGOSFSQGTTSAGHLAGEPAPPPPPPPR 227
DB 256 ---AERONEGOGVEIN-----MATSGN-GQGSTTMDH----- 285
QY 228 EPFAPSLGSAFHLDPAPPAAMAAALYSSSTLPAPPRGSPPLAAPQGSPTKLOGGSAP 287
DB 286 -----ETASVLSSTSHAPRLTSLG-----TIVENYVS-- 316
QY 288 EGATYAPRGSPPKOSPSRLAKSYSTSPINIVVSSAGLSPIRVTSPPTVOSTISSPPIH 347
DB 317 ---LISMGTGHDKMDMSLTMLAMSSSQSCISMRSQGLPLLI----- 356
QY 348 QLSSTIGYATVLSPTKRLVHASEQYSKHSQELVYATATLQPPSLAAGSRAVSSQHG--H 405
DB 357 -----QLHGNKDS-----VLGNSRSGSKARAKAS--AALHNITH 391
QY 406 LGPE-----LRALOSPEHHIDPIYEDRVYOKPPKRSLSQSGDPLPAHNGTYRT 455
DB 392 SQPDKRGRREIVLHLLEQIRAYCETCWEQDAHEPGM---DQDKNMP----- 438

```



```

QY 456 STAPSSGVDVLPQRTGSGHGFQNAATAFQASVAGPASYAGPASYADFPYQLOQCPSEVSE 515
D 439 -----APVEH-----QICPAV-----449
QY 516 YKSGPALPPEGLIARSPSTDSIOKDP-REFGRDELPVYIOMLO-----HOE 563
D 450 -----CVLMKLSDEEHRHAMNMLG-----GLQAIABLLVDCEMYGLTNDHYS 493
QY 564 PSVOSNAAYALHLCFEDNKAIEI-RRGGIQLVLDLIDHRTVEYHSAGCALRNLVYG 622
D 494 IITRRYAGMALTYLTEDVDANKATLCSMKCMALVYQKSESEDIQOYIASVLRNLWS- 552
QY 623 KANDNKIALKCGGIPALVRLIKTTDLIRLVTGVIMNLSS-CDALKMP1-1QDAL 679
D 553 RADVNSKKTREYGSVYALMECALVEYKKESTLSKVSALMNLISAHCTENKADICAVDAL 612
QY 680 AVLNNVLIHSGMENSPLDDDKIQLHSQVLRNATGCLRNSS-AGEEARRMRREC 736
D 613 AFLVGTILTY-----RSQTNLILIESGGILIRNVSILATINEDHROILREN 658
QY 737 DGLTDLALVYIOSALSGSEISDKEVCVILRNLISYRLAETSQGOHMGTDLDGLCG 796
D 659 NCLQTLILHKKSHSL-----TIVSNACGTLMNLSAR-----689
QY 797 EANGDAESSGCGKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKKK 856
D 690 --NPKDQA-----LWDMGAVS-MLKNLH 711
QY 857 CSNPDTEGAAGALONLAA-----GSMKWSYIRAVKREKGLPIVELL- 901
D 712 SKRIMAMGSAALRLNLMANRPKAYDANIMSGSSLPSIH-----VKKQALAEIADAOH 767
QY 902 -----RIDNDRYVCAVATALRNALDVRKELIGRYAMRDVHRLPGGNSNNTASKAM 955
D 768 LSEFEDIDN-----LSPKASHRSKORHKQSLGYDF-----DNNRHDNNSDNF 813
QY 956 SDDTVAVACCTLHEVITKKNENAKALPDAGIGIKLVGISKSG-----D 999
D 814 NGNMNMTVLSPLMTTYLTPSSSSSGSLDSRSREKRLERERIGLGNYHPATENPGTSS 873
QY 1000 KHSPPKVAASQVLYNSMGQVRLDRLSLYK---KQMSQVHFVYASSSTIERDQRPYSSST 1056
D 874 KRLQJSTTAQIAKAMEEVAISHTQEDRSSGSTTELHCY---TDERNALRSSAAHT 929
QY 1057 PS--ISPPRVSPNKRASAPASPREMISLKERKTDYECTGSNATYHGAKGHTSRKDAT 1114
D 930 HSNMTYFTKSENENRTCSMPYAKLE---YKRSNDLSNVSSSDGYKRGQ---MKPSIE 983
QY 1115 AQMTGISTLYRNSYGAPAEIDIKHNOVSADQVPEPSRKDIETIYQPNSTRNDE 1169
D 984 SYSEDESKF-CSYGOYPADLAKHKHSANHMD---NDGEIDTPIYSLKYSDE 1033

```

```

ADDRESS: Banner & Wilcoff, Ltd.
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/452,655B
FILING DATE: 25-MAY-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/289,548
FILING DATE: 12-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/741,940
FILING DATE: 08-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107,49964
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9299
TELEFAX: 202-508-9100
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2843 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-452-655B-2

Query Match 3.0%; Score 194; DB 1; Length 2843;
Best Local Similarity 18.5%; Pred. No. 5.6e-05;
Matches 232; Conservative 164; Mismatches 428; Indels 432; Gaps 49;

QY 23 SASSETSLSP-----GLNTSNDGSETE-TTSAILIASYKQE-----LQ 61
D 102 SVSSHSRGCSFPMGSPFRGTFVNGSRESTGYLBELERSLIADLDKEKEKDWYAAQ 161
QY 62 FERLIRELEA-----EROIVASOLE-----RCKLGETSGMSMSAEPQFOMOS 106
D 162 LQNLTKRIDSLEPNFSLQDMDTRROLEYEARQIRAMEQLQTCQDMKRAORRIARI 221
QY 107 QDQKDIEDLTTGLELVDSICIRLSQESGILDPDYSTGERPSLLSQSALQLNSKPSGF 166
D 222 QOIERDI-----LRIRLOLOS-----QATEAERS-----QNKHETGSH 255
QY 167 QYPASVHNOTLALGETTPSOLPARGTQARATGQSFQGTTSRAGHLAAGERAPPPPPP 226
D 256 D--AERONEGQVGEIN-----MATSGN-QGGSTTRMDH-----286
QY 227 REPAPSLGSAFHLPDAPPAALAAALYSSSTLPAPPGSGSPLAAPGSGPTKLQRGSA 286
D 287 -----ETASVLSSTSTHAPRLTSHLG-----TKVEMYSS- 317
QY 287 PEGATYAAPRGSSPKQSPSLAKSYSTSSPINIVYSSAGLSPIRVTSPPVQSTISSPT 346
D 318 -----LISMGLTHDKDMSRTLLAMSSSDSCISMROSGCLPLLI-----357
QY 347 HOLSTIGTAVTISPTKRLVHASFOYGHSELYATATLQRPGLSAGSASYSOQG- 404
D 358 -----QLHGNDKDS-----VILGNSRGKEKARARS-AALHNI 391
QY 405 HUGPE-----LRLAQSPENHIDPIYEDRYOKRPPMRSLSQSGDLPAPAHGTGR 454
D 392 HSOPDDKRGREINVLHLEDIRAYVCEIWCWMCNHEPQM-----DQKXNMP-----459

```

```

OY 455 TSTAASSPGVDSPLOFTQSOHGQONAAATFORASAYAGPASNADPYROLQYCPSEVES 514
DB 440 -----APVEH-----QICPAV-- 450
OY 515 PYSKSGPALPEPGLIARSPSIDSTOKDP-REFGNRDEPLEVEIOMQ-----HQ 562
DB 451 -----CYLMKISFDEEHRHAMNELG-----GLQAIABLLQVDCEMYGLTMDHY 493
OY 563 FPSVOSNAAYLOHCFEGDNKIKAEI--RROGGIOLLDVLDHRTMEVHRSGALRLNLY 621
DB 494 SITLRYGAMALNTLTFEDVANAKATLCSMKCMRALVAQLSESDIQQYIYASVLRNLSW 553
OY 622 GKANDNKAALKNGGIPALVRLKTTDEIRRELYGVLMNLS--CDALKMPI--IODA 678
DB 554 -RADVNSKKTREYGVSAALMECALVEYKKESTLSKVSALMNLNLANHCTENKADICAVDGA 612
OY 679 LAVLTNAVITPHSGWENSPLODDRKIQLHSSQVIRNATGCLRNVS--AGEEARRRRE 735
DB 613 LAFVGLTLY-----RSQTNLTALIESGGGILRNYSGLIATNEDRHQILRE 658
OY 736 CDGLDALLVYIOSALGSSEIDSKTVENCVCILNLSVRLAETSOQOHMGTDLDGLLC 795
DB 659 NNCLQTLQHLKSHL-----TIVSNAGCTLMNLSR----- 690
OY 796 GEANGKDAESSGCGWKKKKKKKSDODMGVPLPDCAEPKGIOMLWHPISVKEPYLLLS 855
DB 691 ---NKDOEA-----LMDMGVNS--MLANLI 711
OY 856 ECSPDITLEGAGALONTAA-----GSMKWSYIIRANVAREKGLPILVELL 901
DB 712 HSKHKMIAMGSAALRLNMANRPAKYKDANIMSPGSLPSLH-----YKQKALAEALDAQ 767
OY 902 -----HIDNDRVYCAVATLARNALVNRKELIGKAYAMDVLHRLPGCNSNNTASKA 954
DB 768 HILSEFDNIDN-----ISPKASHRSKORHKOSLYGDIYF-----DTRHDDNNSDN 813
OY 955 MSDDTVAVCCTLHEVITRKNMENAKALRDAGIEKLVGISKSG----- 998
DB 814 FNTGMATVLPSPITVLPSSSSSRGSLDSRSEKDRSLERERIGIGNYHPATENPOTS 873
OY 999 DKHSKRYKAAQVINSWQTRDLRLSLK---KQMSQYHRYVASSSTIEROROPYSSSR 1055
DB 874 SKRGLOISTTAAQIAKVAEEVSAIHTSQEDRSSGSGTTELHCY---DTERNALRRSSAH 929
OY 1056 TFS--ISPVRSVNNRNSAPASPREMISLKERKTDYECTGSNATYHAKGEHRSRKDAM 1113
DB 930 THSNYNTFKSNSNRKTCGSMYAKLE---YKRSSNDLSNYSVSSDGYCKRQ---MKPSI 983
OY 1114 TAQNTGISTLYRNSYGAPAEIDIKHQAQVAPVPOEPRSKDYETYQPFONSTRNATDE 1169
DB 984 ESYSEDESKF-CSYGOYPADIAHRIHSANHMD---NDGELDTPIYNSLKYSDE 1034

```

```

RESULT 6
US-08-452-655B-7
; Sequence 7, Application US/08452655B
; Patent No. 5783666
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 102
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Wilcoff, Ltd.

```

```

; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452,655B
; FILING DATE: 25-MAY-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/289,548
; FILING DATE: 12-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/741,940
; FILING DATE: 08-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107,49964
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2843 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ANTI-SENSE: NO
; US-08-452-655B-7

Query Match 3.0%; Score 194; DB 1; Length 2843;
Best Local Similarity 18.5%; Pred. No. 5.6e-05;
Matches 232; Conservative 164; Mismatches 428; Indels 432; Gaps 49;

OY 23 SASKTSLSLP-----GLTNSGDSGSETE-TTSAIILASVKEQF-----LQ 61
DB 102 SVSRSRGECSPVPMGSPFRPGFVNGRSESTGYLELEKERSILLDLDEEEKDMYAAQ 161
OY 62 FERLTRELEA-----EROIVASOLE-----RCKIGSETGSMSSSAEEOFQMS 106
DB 162 LQNLTRIDSLPTENFSIQTDTRQLEYEARQIVAAEEOGLTQDEKRAQRIARI 221
OY 107 QDQKDIIEDELTTGLBLVDSCIRSLQESGILDPDQYSTGERPSLLSQSALQUNSKPSF 166
DB 222 QQIEKDI-----LRIQLQLQS-----QATEAERS-----QNKHETGSH 255
OY 167 QYPAHSNQTALGETTTSQQLPARGTQARATGQSPSGQTSRAGHLAGPEAPPPPPP 226
DB 256 D---AERQNGGQGVGIN-----MATSGN--GQGSTTRMH----- 286
OY 227 REPFARSLGSAFHLPPAPRAAAAALYSSSTLPAPPRGSGPLAAPGSGPTKLRGSA 286
DB 287 -----ETASVLSSTSHSAPRLTSLG-----TKVENYVS-- 317
OY 287 PEGATYAAPRGSSPKQSPSLAKSVSTSSPINIVYSSAGLSPIRYTSPPTVOSTISSPI 346
DB 318 -----LLSMGLGTHDKDDMSRTLLAMSSQDSCISMROSCLPLLI----- 357
OY 347 HQLSSTIGTATISPTKRLVHASEQYSKSHQELIYATATQIRGSLAAGRASYSOHHG-- 404
DB 358 -----QLHGNKDS-----VLNRSRGSKEARRAS--AALHNIT 391
OY 405 HIGPE-----LRAQSPENHIDPIYEDRVYQKPMRSLSQSGPLPPAHTGYR 454

```

```

Db 392 HSPDDBKRGREIRVHLLEQIAYCETCWEQEAHEPGN-----DQDKMP----- 439
QY 455 TSTAPSSPGVDVPLQRTGSHGPONAAAATFORASVAGPASNVADPYRQLOCPSVES 514
Db 440 -----APVEH-----QICPAV----- 450
QY 515 PYKSGPALPEPTLARSPSIDIQDP-REFGWRPELEPEVIOUO-----HQ 562
Db 451 -----CVLMKLSFDEHHRHAMELG-----GQALAEILOVCEMYGLTNDHY 493
QY 563 FPSVOSNAAYLOHLCFEGDNKIKAEI-RRQGGTQLLVLDLDHMTVEVHSAGCALINLVY 621
Db 494 STLRIRAGALINLTFPGDVANAKATCSMKGCRRALVAQLKSESDLOQVIAVLNLSM 553
QY 622 GRANDNKIALKNCGGIPALVRLRLTDLRELYTGVLMNLS-CDALKMP1--TDA 678
Db 554 -RADVNSKTLRFVGSVKALMECALFEVKKESTLSVLSALMNASCTENKADICAVDA 612
QY 679 LAVLTAAVITPHSGWENSPLODRKIQULHSOVLNATGCLRVSS--AGEBARRRMR 735
Db 613 LAELVGTLY-----RSQTLALIESGGILIRNVSSLIATNEHDHRIKRE 658
QY 736 CDGLTALLVYISALGSSEIDSKTEWNCVILRNTSYRLAETSGQMGHTDELGLJC 795
Db 659 NNCLOTLLOHLKSHSL-----TYSNACGTMNLSAR----- 690
QY 796 GEANGDAESSGCGWKKKKKKKQSDQWDGVLPLDCAEPPKGIOMLMHPSYKPYLTLS 855
Db 691 ---NPKDOEA-----LMDGAVS-MLKNLI 711
QY 856 ECSNPITLBEAGALONIAA-----GSMKSVYIRAAVREKGLPIVELL 901
Db 712 HSKHKMTAMGSAALNLANRPAAKKNANIMSPGSSLSLH---VRQKALEAELDKQ 767
QY 902 -----RIDNDRYCAVATATLNNALDYRNKELIGKYAMRDLVHRLPGGNSNNATSKA 954
Db 768 HLETFDNDIDN-----LSPKASHRSKQHRKOSLYGDYF-----DTNKHDDRSN 813
QY 955 MSDDTYAVCCTLHEVITTKMENAKALRDAGIEKLVGISKSG----- 998
Db 814 FNTGNMTVLSPLYNTVTLSPSSSSSGSLDSSREKDRSLEREIGLIGNVHPATENPGTS 873
QY 999 DKSPRYVAASOVLNMMQYRDLASLYK---KDGMSQYHFASSSTIRDRQRPSSSR 1055
Db 874 SKGGLDITTAQIAKAYMEVSAIHTSOEDRSSGSTTELCY---TDRNALRRSSAH 929
QY 1056 TPS--ISPRVSPNNNSASAPASPREMISLKERKTDYECTGSNATYHGAKEHTSKDAM 1113
Db 930 THSNNTNFKSENSNRTCSMPYAKLE---YKRSNDLSLNVSSSDGCGKRGQ---MKPSI 983
QY 1114 TAONTGISTILYRNSYCAPADIKHNOVSAQVPQEPFSKRDYETYOFPONSTRNDE 1169
Db 984 ESYSEDESKF-CSYGQYPADLAHKIHSANHMD---NDGELDPINYSLSYSDR 1034

RESULT 7
US-08-450-582-2
; Sequence 2, Application US/08450582
; Patent No. 6114124
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARI
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YOSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 102

```

```

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Wilcoff, Ltd.
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/450,582
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/452,655
; FILING DATE: 25-MAY-1995
; APPLICATION NUMBER: US 08/289,548
; FILING DATE: 12-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/741,940
; FILING DATE: 08-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107,49964
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2843 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-450-582-2

Query Match 3.0%; Score 194; DB 4; Length 2843;
Best local similarity 18.5%; Pred. No. 5.6e-05;
Matches 232; Conservative 164; Mismatches 428; Indels 432; Gaps 49;

QY 23 SASERTSLSP-----GLNTSNGGSETE-TTSAILASYKEQ-----LQ 61
Db 102 SVSSNSGCSVPVPMGSPFRGFGVNGSRSTGYLELEKERSLLADLDEKEKDWYAAQ 161
QY 62 FERLTRELEA-----EROIVASOLE-----RCKLSGETGSMSSMSAEQFQWOS 106
Db 162 LQNLTKRIDSPLTENFSLQDMTRROLEVEAROIRVAMEBOLCTCODMERAQRIARI 221
QY 107 QDGKQIDDELTTGELVDSCIRSLQESGLDPQDYSGERPSLLSSALQNLNKKPGSF 166
Db 222 QOIEKDI-----LRIRQLLOS-----QATEAERS-----QNHETGSH 255
QY 167 QYPASYSNQTLAGETTPSQLPARGTQARATGQSFQGTTSRAGHLAGEPPAPPPPP 226
Db 256 D--AERQNEGQVGEIN-----MATSGN-GQGSTTRMDH----- 286
QY 227 REPRAPSLGSAFHLPDAPPAANAALYSSSTLPAPRPGSPPLAAPGSGPTKLQRGGA 286
Db 287 -----ETASVLSSTSTHAPRLTSHLG-----TVEWYYS- 317
QY 287 PEGATYAAPRGSSPKQSPRLAKSYSTSPINIVSSAGISPIRVTSPTVQSTISSPI 346
Db 318 -----LISMCTHDKDMSKTLTAMSSSDSCISMROSGLPLIT----- 357
QY 347 HOLSTIGTATLSPTRKVLVHASDOYKSHQELATATLQRPGLSAGSRASYSQHG-- 404
Db 358 -----QLHNGNDKDS-----VLGNSRSGKEARANAAS-AALHNII 391
QY 405 HLGPE-----LRLQSPENHIIDPIYEDRYQKPPKRSLSQSGDPLPRAHGTTR 454

```

Db	392	HSQDDDKRGRREIRVHLHLEQIRAYCETCMEQEAHEPGH----	DDDKRPM-----	439
OY	445	TSTAASSEPVDSVPLRQTSQHGPONMAAATFORASTYAAGPASNADYPRLOJYCESVES	514	
Db	440	-----APVEH-----	QICPAV----	450
OY	515	PKYSGGALPEPEGLIARSPSIDSIQKDP-REFGWRPELPEVIOMQ-----	HQ 562	
Db	451	-----CVLMKISFDEBHRMANEL-----	GLQALAEILLQVDCEMVGLTNDY 493	
OY	563	FPVSQNAAYLOHICFGBDKIKAEI-RRQGGIQLVLDLIDHRETEVHRSAGCALRNEY	621	
Db	494	SITLRIRAGALNLTUTTGVDANKATLCOSKMGCBALYALQKSESDLOQVIA SVLRNLSM	553	
OY	622	GRANDDKIALKNGGIPALVRLKRTDLEIRELVGYLWNLSS-CDALKMPI--TODA	678	
Db	554	-RADVNSKKTLEVGSVYKALMECALVEYKRESTLKVIALMNLSTACTENKADICAVDA	612	
OY	679	LAVLTNAVIIIPHSWENSPLODDRKIQOLHSOVLNRATGCLRVS--AGEARRRME	735	
Db	613	LAFVLGILTY-----RSQTNFLAIEGGGLRVSLLIATNEBHROIRE	658	
OY	736	CDGLTDLALVYIOGALGSEIDSKYVENCICILRNYSYRLAETSGQMGHGDDELQILC	795	
Db	659	NNCIQTLQHLKHSL-----TYSNACGTLMNISAR	690	
OY	796	GEANGDAESSGCGKRRKKRSODOWDVGPRPDCAEPPKGIOMLNHSIYKPYLLLS	855	
Db	691	---NPKQGEA-----	LMDGAVS-MLKNLI 711	
OY	856	ECSNPDTLEGAGALQMLA-----	GSKMSVYIRAAVREKKEGILPIVELL 901	
Db	712	HSKRMTAMGSAALNRIMANRPAYKDANIMSGSLPSLH-----VRQKALEAFELQ	767	
OY	902	-----RIONDVYCAVANATLRMALDVANKTELIGVYAARDLVHRLPGGNSNNTASKA	954	
Db	768	HLSTFEDNID-----LSPRKASHRSQRRKOSLYGDIYF-----	DTNRHDKRDN 813	
OY	955	MSDDTVAVACCTLEHVTTKMMENAKALRDAGIEKLVIJSKSG-----	998	
Db	814	FNTGNMTVLSYLTNTVLPSSSSSRGSLDSREBKDRSLERENGIGLGNHPTENPGS	873	
OY	999	KHSPEKYVKAASOVLNSMMQYRDLRSLYK---KDGMSQIHFNVASSSTIERDORPYSSR	1055	
Db	874	SKRLOISTTMAQITAKWEEVSAIHTQOEDRSSGSTTEILHCY-----TDERNALARSSAAH	929	
OY	1056	TPS---ISPVRSPNNRBSASAPASPREMISLKEKRTYECGSGNATYHGAKEHTSKDKM	1111	
Db	930	THSMTYNFTKSENENRCSMPYALE---YKRSNOSLNVSSSDGIGRKGQ---MKPSI	983	
OY	1114	TAONTGISTLYRNSYGAPAEIDIKINQVSAQPVQEPERKQETYYOFPONSTRNRYDE	1169	
Db	984	ESYSEDDDESK-CSTGYQYPADLAKHISAHNMD-----NDGEIDTPIINTSLKTSDE	1034	

RESULT 8  
 US-08-450-582-7  
 : Sequence 7, Application US/08450582  
 : Patent No. 6114124  
 :  
 : GENERAL INFORMATION:  
 :  
 : APPLICANT: ALBERTSEN, HANS  
 : APPLICANT: ANAND, RAKESH  
 : APPLICANT: CARLSON, MARY  
 : APPLICANT: GRODEN, JOANNA  
 : APPLICANT: HEDGE, PHILIP J.  
 : APPLICANT: JOSLYN, GEOF.  
 : APPLICANT: KINZLER, KENNETH  
 : APPLICANT: MARKHAM, ALEXANDER F.  
 : APPLICANT: NAKAMURA, YUSUKE  
 : APPLICANT: THLIVERTIS, ANDREW  
 :  
 : TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC  
 : TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS

```

NUMBER OF SEQUENCES: 102
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Witcoff, Ltd.
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/450,582
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/452,655
FILING DATE: 25-MAY-1995
APPLICATION NUMBER: US 08/289,548
FILING DATE: 12-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/741,940
FILING DATE: 08-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107,49964
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2843 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEetical: YES
ANTI-SENSE: NO
US-08-450-582-7

Query Match          3.0%: Score 194; DB 4; Length 2843;
Best Local Similarity 18.5%: Pred. No. 5,6e-05;
Matches 233; Conservative 164; Mismatches 428; Indels 432; Gaps 49.

QY      23 SASERTSLSP-----GLTNSGDSGETE-TTSAILIAYKEDE-----IQ 61
       | : : : | | | | | | | | | | | | | | | | | | | | | |
DB      102 SVSRSGECSPVDMGSPRRGFNGSRESTGYLIELEKERELLADLDKEKEDWYAAO 161
       | : : : | | | | | | | | | | | | | | | | | | | | | |
QY      62 FERITRELEA-----EROIVASOLE-----RCKLGSETSMSSMSAEBOFOMOS 106
       | : : : | | | | | | | | | | | | | | | | | | | | | |
DB      162 LQNTFKRIDSLPLENFSLQTDTRQOLEYEAKQIRVAMEEOQLTCODMEKKARQRIARI 221
       | : : : | | | | | | | | | | | | | | | | | | | | | |
QY      107 QDGQKDEDELDTGLIELVDSCIRLSQESILDPDQYSTGERPSLLISALQNLKREGEF 166
       | : : : | | | | | | | | | | | | | | | | | | | | | |
DB      222 QQIKNDI-----LRIRQLLOS-----QATAEKSS-----QNKHETGSH 255
       | : : : | | | | | | | | | | | | | | | | | | | | | |
QY      167 QYPASYSNOTLALGETTPSOULPARGTQARATGGSFGSTTSRKAGHLAGEPRAPRRPPRR 226
       | : : : | | | | | | | | | | | | | | | | | | | | | |
DB      256 D---AERQNQGCGEGEIN-----MATSGN-GGGSTTRMDH----- 286
       | : : : | | | | | | | | | | | | | | | | | | | | | |
QY      227 REPAPRLGSAFHLPAPAPMAAAALYYSSSTLPAPPGGSPILAPOGGSPTRKLDRGCSA 286
       | : : : | | | | | | | | | | | | | | | | | | | | | |
DB      287 -----ETASVLSSSTHSAPRLTLHLG-----TKEMWYS- 317
       | : : : | | | | | | | | | | | | | | | | | | | | | |
QY      287 PEGATVAAPRGSSPKOSPRLAKSVSTSSPINIIVVSAGLSPIRYTSPPTVQSTISSSPI 346
       | : : : | | | | | | | | | | | | | | | | | | | | | |
DB      318 ----LLSMIGTHDKDDMSRTLLAMSSQDSCISMROGCPLLI----- 357
       | : : : | | | | | | | | | | | | | | | | | | | | | |
QY      347 HQLSSTTGTAITLSPTRKRLVHASQYSKISQELVATATLQRPGLSLAAGSNASYSQQHG-- 404

```



Db	392	HSOPDDKRGREIVLHLLQD	IRAYCTCMEQDANEHFG	----	DDDKNMP	-----	433		
OY	445	TSTPSSSGVDVSVLQDTSG	OHGFOHMAAATFORASTA	AGPASNUTADPTROLOJCS	VSYS	514			
Db	440	-----	-AVEH	-----		450			
OY	515	PYSKGPAPEPGLARSPSID	SIOKDP	-REPGRDPELPEVLOMO	----	HO	562		
Db	451	-----	-CVLMKLSFDEHRIANMELG	----	GLQATIELLOQDCEMTGLT	INDHY	493		
OY	563	FPVSQSNAAVLOHLCFGDN	KIKAEI	-RROGQIQLVDLIDHMT	EVNHSRAGCALRNLY	621			
Db	494	SITLRRYAGMALITMLTGE	VDVANKATLCMSKGCRA	LYAQLKSESEDIQOYASV	LENLSM	553			
OY	622	GKANDDKIALKNGGIPALV	RLRLKTTDEIETELVGLV	MINSS	-CDALKMP	--IQDA	678		
Db	554	-RAVYNSKRTLREVGSYKA	LMECALVEKKESTLKY	SVLSALMINSACTENK	ADICAVDA	612			
OY	679	LAVLTNAVITPHSGWENS	PIODDKIDHSSOVLRYNAT	GCLRNVSS	----	AGEARRRMR	735		
Db	613	IAFLVGLITY	-----	RSQTNLAIITESGGIL	RAVSSLIATNE	DHROITLE	658		
OY	736	CDGLTDALLYIOSALGS	SEIDSKTYENCYILRN	LSYRLAETSOQHNGTDEL	DGLLC	795			
Db	659	NNCLOTLOHLKHS	L-----	TIYSNACGTLMIN	SAR	-----	690		
OY	796	GEANGRADESSGCGKKK	KKKKKODDOWDVGPR	PDCAEPKGIQIMLH	HSYKPYITLIS	855			
Db	691	-----	-NPKOEA	-----	-LMDGAVS	-MLKNLI	711		
OY	856	ECSNPTLEGAAGALONIA	A-----	GSWKSVYIRAVRREK	GPLIVELL	901			
Db	712	HSRKIMIAMCSAALRL	UDMANRPKAKYDANIM	SGSSLSPLH	----	YRKOKALEALDQ	767		
OY	902	-----	-RIDNRYVCNAVATALRN	MALDVBNKELIGKYAM	RDLVHRLPGNNS	NNTASKA	954		
Db	768	HLSETPENIDN	-----	LSPKASHRSQKHQSL	GYDYP	-----	DTNRHDDRSN	813	
OY	955	MSDPTAVACCTLHEVIT	KTMENAKALRDAGLE	KLVGISKSG	-----		998		
Db	814	FNICNMIVLSEPLYTTL	VLPSSSSRGSLDSREK	DSLEREGICLGNYH	PATENPGTS	873			
OY	999	DKHSPPKVAASOVLNS	MOQYRDLRLSYK	----	KDGMSQYHFVASS	STIEDRORPSSR	1055		
Db	874	SKRGLQSTTAQAQAKM	EBVSAIHNSQEDRSSG	STTELHCV	----	TDERNMLRSSAH	929		
OY	1056	TPS	-----	ISPVRSPPNRSASAP	ASPREMISLKEKRTD	ECTGSNATYHAKAGE	HTSKDAM	1111	
Db	930	THSNTVMTFTESEN	NPTCSMPYAKLE	----	YKRSSNLSLVS	SSSDGYGRQ	-----	MKPSI	983
OY	1114	TAQNTGISTLYRNSYGA	PADIDHNOVSQAPVQ	EFERSKDYETUOFO	QSTRNRYDE	1169			
Db	984	ESISEDESKF	-CSTGYQPADLHLK	HTSHANMD	-----	NDGELDPTIN	SLKYSD	1034	

RESULT 10  
US-09-003-687A-7  
; Sequence 7, Application US/09003687A

? GENERAL INFORMATION:  
 ? APPLICANT: Barker, Nick  
 ? APPLICANT: Clevers, Hans  
 ? APPLICANT: Korinek, Vladimír  
 ? APPLICANT: Morin, Patrice  
 ? APPLICANT: Kinzler, Kenneth  
 ? APPLICANT: Vogelstein, Bert  
 ? APPLICANT: Sparks, Andrew  
 ? TITLE OF INVENTION: Beta Catenin, TCF-4, and APC  
 ? TITLE OF INVENTION: Interact to Prevent Cancer  
 ? NUMBER OF SPOUNCES: 11  
 ? CORRESPONDENCE ADDRESS:  
 ? ADDRESSEE: Banner & Witcoff, Ltd.  
 ? STREET: 1001 G Street, N.W.

```

1      CITY: Washington
2      STATE: DC
3      COUNTRY: USA
4      ZIP: 20001
5      COMPUTER READABLE FORM:
6      MEDIUM TYPE: Diskette
7      COMPUTER: IBM Compatible
8      OPERATING SYSTEM: DOS
9      SOFTWARE: FASTSEQ for Windows Version 2.0
10     CURRENT APPLICATION DATA:
11     APPLICATION NUMBER: US/09/003,687A
12     FILING DATE:
13     CLASSIFICATION:
14     PRIOR APPLICATION DATA:
15     APPLICATION NUMBER: 08/821,355
16     FILING DATE: 20-MAR-1997
17     ATTORNEY/AGENT INFORMATION:
18     NAME: Kagan, Sarah A
19     REGISTRATION NUMBER: 32,145
20     REFERENCE/DOCKET NUMBER: 1107.05064
21     TELECOMMUNICATION INFORMATION:
22     TELEPHONE: 202-508-9100
23     TELEFAX: 202-508-9299
24     TELEX: 97430 BMB UT
25     INFORMATION FOR SEQ ID NO: 7:
26     SEQUENCE CHARACTERISTICS:
27     LENGTH: 2973 amino acids
28     TYPE: amino acid
29     STRANDEDNESS: single
30     TOPOLOGY: linear
31     MOLECULE TYPE: NO. 5998600e
32     US-09-003-687A-7

```

Query Match	3.0%	Score 194	DB 2	Length 2973
Best Local Similarity	18.5%	Pred. No. 6e-05		
Matches 222	Conservative 164	Mismatches 428	Indels 432	Gaps 49
QY	23	SASEKTSLSLP	-----GLNTSNDGSETE--TTSAILAYKEOE-----LQ	61
Db	102	SVSSRSQECSPVMGSPFRGVNSRSESTGYLELEKEKRLTLADDKDEKEKDWYAAQ	161	
QY	62	PERITRLEA	-----EROIVASOLE-----RCKLSGETSGMSMSMAEQQOMOS	106
Db	162	LQNTKRIKISLPTEENESLQTMTRQOLEYBARQIRVAMEBQLQTCODMKERRAQRIRAI	221	
QY	107	QDQCKIDEDELTGLVDSICRSLOESGILDPDQSYNGEPSPILSSQALQMSKEGFSF	166	
Db	222	QOIKEDI	-----LRITQLLOS-----QATFARRS-----QMKHETGSH	255
QY	167	QYPASVHSMQTALGETTPRSQLPARGTARATGOSFSOGITTSRAGILAGEPAPPPPPP	226	
Db	256	D---AEKQNEQGVGEIN	-----MATSGN--GQGSTTRMDH-----	286
QY	227	REFPAPSLGSAFHLPPAPPAALALYYSSSTLPAPPRGSGPLAARQSGSPTLQRGSA	286	
Db	287	-----	-----ETAVALSSSSTHSAPRLTSHLG-----TVEVWYS--	317
QY	287	PEGATYAAPRGSSPKOSPRLKAYSTSTSPINIVWSSAGSPIHVTPTVGTSTISSPI	346	
Db	318	-----LLSMIGHDKDDMSKRTLLANSSQDSCISMROSGCLPLII	357	
QY	347	HOLSTTGTATLSPTKRLVHASEQYSKHSOELATATLQRPGLSALAGSRASYSSQHG--	404	
Db	358	-----	-----QLLHGNDKS-----VLLGNSRSGSKEARARS--AALHNII	391
QY	405	HLQPE	-----LRALQSEHHIIDITYEDRYAQKPPMRSLSSQSGDPLPAAHGTTR	454
Db	392	HSQPDKKRGRREIRVLIHLEQIRAYCECTCWMQEAHEPGM-----DQDKNMP	439	
QY	455	TSPAPSPGVDSVPLQRTGSGHGQNAANAATFGQASVTAAGPASNVDAPYKQLOLQCYSES	514	
Db	440	-----APVEH	-----QICGAV--	450

```

OY 515 PYSKSGPALPECTIARSPISIDOKP-REFGWRDEPEVYIOMQ-----HQ 562
Db 451 -----CVLMKLSFDEEHRHANNELG-----GLQAIALLQVDCMYGLTNDHY 493
OY 563 FFSVQSNAAVYLOHLCFGDNKIKAEI-RRGGIQLLVYDLDDHMTVEVHRSACALNIVY 621
Db 494 SITLRYAGMALNTLTGGVANKATLCGSMGCMALVAQLKSESEDLQOVYASVLRNLSW 553
OY 622 GRANDNKIALKNCIGIPALVRLRKTDTLEIRELYGVLMNLS-CDALKMPI--IQA 678
Db 554 -RADVNSKTLREVGSYKALMECALEVKKESTLKVSLALMNLASHCTEKADICAVDA 612
OY 679 LAVLTNAVIIPHSGWENSPLODDRKTQLHSSOYLARNATGCLRNVS--AGEBARRRMR 735
Db 613 LAVLVGLTY-----RSQTNLTATIESGGGILRNVSLLIATFNEDHQILRE 658
OY 736 CDGLTDALLVYIOSALGSSEIDSKTYENCVCILRNLSYRLAETSGOCHMGTELDGILL 795
Db 659 NNCLQTLLOHLKSHSL-----TIVSNACGTLNLSAR-----690
OY 796 GEANGKDAESSGCGMKKKKKKSSQDQMDVGPLPDCAEPKGIOMLMHPSIVKPYLTLLS 855
Db 691 ---NPKQEA-----LMDGAVS-MLKNI 711
OY 856 ECSNPDTLEGAGALONLAA-----GSKWVSYIIRAAVREKGLPIVELL 901
Db 712 HSKRMKIAMGSAALRNLMANRPAYKADANIMSPGSLPSLH--VRKQKALEELDAQ 767
OY 902 -----RIDNDRYVCAVATALRNALDVNRKELIGKYAMRDVLRHLPGGNSNNTASKA 954
Db 768 HLESTFNDIN-----LSPKASHRSQRKHQSLYGDYVF-----DTRNRHDNRSDN 813
OY 955 MSDDTVAVOCCTLEHVTIKNMENAKALRDAGIEKLVGISKSG-----998
Db 814 FNTGNTVLSYLTLYTPSSSSSRGLDSRSEKSLERERIGICNTHPATENDGTS 873
OY 999 DKHSPIYKKAQOVLNSMWQYRDLRLSYK--KDWMSQYHEVASSSTIERDRQRPYSR 1055
Db 874 SKRGQISTTAQIAJAKMEVSAIHTSOEDSSSGSTELHCV-----IDERNALRRSSAH 929
OY 1056 TPS--ISFVRSPPNRSASAPASPREMISLKERKTDYECTGSNATYICANGEHTRKDM 1113
Db 930 THSTYNTTKSENENRPTCSMPYAKLE--YKRSSNDLSINSVSSDYGKRGQ--MKPSI 983
OY 1114 TAONTGISTLRNSYGAAPEDIKHNQVSAQVPEPSRKQDETQOPONSRYNDE 1169
Db 984 ESYSEDESKF-CSTGOYPADLAHKHSANHMD--NDGELDTPIYNSLKYSD 1034

```

RESULT 11

```

US-09-136-605-7
Sequence 7, Application US/09136605A
Patent No. 6140052
GENERAL INFORMATION:
APPLICANT: He, Tong-Chuan
APPLICANT: Kinsler, Kenneth
APPLICANT: Vogelstein, Bert
TITLE OF INVENTION: Beta Catenin, TCF-4, and APC Interact to
FILE REFERENCE: 1107.75741
CURRENT APPLICATION NUMBER: US/09/136.605A
EARLIER FILING DATE: 1998-08-20
EARLIER APPLICATION NUMBER: 08/821,355
EARLIER FILING DATE: 1997-03-20
EARLIER APPLICATION NUMBER: 09/003,687
NUMBER OF SEQ ID NOS: 28
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 7
LENGTH: 2973
TYPE: PRT
ORGANISM: Homo sapiens

```

US-09-136-605-7

Query Match 3.0%; Score 194; DB 4; Length 2973;  
 Best Local Similarity 18.5%; Pred. No. 6e-05;  
 Matches 232; Conservative 164; Mismatches 428; Indels 432; Gaps 49;

```

OY 23 SASKTSLSIP-----GLNLSNGDSTF-TTSAIIASVGEQ-----LQ 61
Db 102 SVSSSGECPYPMGSPFRGCVNRSSTGYLELEKERSLIADLDKEKERDWMYAAQ 161
OY 62 FRLTRLELA-----EROIVASOLE-----RCLGSETPGSSMSAEOFOQMS 106
Db 162 LONLTKRIDSLEPTENFSLQDTMTRRQLEYEARQIRVAMEEDLGICQMEKRAQRIARI 221
OY 107 ODGOKIDELFTTGLEIVDSCIRSLQESGILDDPODYSTGERPSLLSOSALOLNSKPEGSF 166
Db 222 QQLENDI-----LRITROLLOS-----QATEAKRS-----QNKHEIGSH 255
OY 167 QYPASYHNOQLALGETTPSQLPANGTQARATGQSFQGTTSRAGHLAGEPAPPPPPP 226
Db 256 D--AERQNEGQVGEIN-----MATSGN-GGGSTTRMDH-----286
OY 227 REPFAPLSASHPDAPPAALAAALYYSSITLPPRPGGSLAPRGGSTPKIQRGSSA 286
Db 287 -----ETASVLSSTHSAAPRLTSHLG-----TKEMVYS-317
OY 287 PEGATVAPRGSSPKOSPSRLAKSYSTSPINIVVSAGLSPIRYTSPPTVOSTISSPI 346
Db 318 ---LISMGLHDKDMSRTLLAMSSSDSCISMRQSCPLLI-----357
OY 347 HOLSTIGTATLSPYKRLVHASQYSHQELVATATLQRPGLSAGSRASYSSQHG--404
Db 358 -----QLHGNKDS-----VLGNSRGSKEARARAS--AALHNI 391
OY 405 HLGPE-----LRAQSPENHIDPIYEDRYVQKPPMRLSQSODDPLPATGTYYR 454
Db 392 HSQPDKRGRREIRYVHLLQIRACETCEWQEAHEPDM--DQKNPDM--439
OY 455 TSTAPSPGVDSVPLQRTSGOHQPQMAAATFORASAYAGPASNADYPRQLOQCPVES 514
Db 440 -----APVH-----QICPAV--450
OY 515 PYSKSGPALPECTIARSPISIDOKP-REFGWRDEPEVYIOMQ-----HQ 562
Db 451 -----CVLMKLSFDEEHRHANNELG-----GLQAIALLQVDCMYGLTNDHY 493
OY 563 FFSVQSNAAVYLOHLCFGDNKIKAEI-RRGGIQLLVYDLDDHMTVEVHRSACALNIVY 621
Db 494 SITLRYAGMALNTLTGGVANKATLCGSMGCMALVAQLKSESEDLQOVYASVLRNLSW 553
OY 622 GRANDNKIALKNCIGIPALVRLRKTDTLEIRELYGVLMNLS-CDALKMPI--IQA 678
Db 554 -RADVNSKTLREVGSYKALMECALEVKKESTLKVSLALMNLASHCTEKADICAVDA 612
OY 679 LAVLTNAVIIPHSGWENSPLODDRKTQLHSSOYLARNATGCLRNVS--AGEBARRRMR 735
Db 613 LAVLVGLTY-----RSQTNLTATIESGGGILRNVSLLIATFNEDHQILRE 658
OY 736 CDGLTDALLVYIOSALGSSEIDSKTYENCVCILRNLSYRLAETSGOCHMGTELDGILL 795
Db 659 NNCLQTLLOHLKSHSL-----TIVSNACGTLNLSAR-----690
OY 796 GEANGKDAESSGCGMKKKKKKSSQDQMDVGPLPDCAEPKGIOMLMHPSIVKPYLTLLS 855
Db 691 ---NPKQEA-----LMDGAVS-MLKNI 711
OY 856 ECSNPDTLEGAGALONLAA-----GSKWVSYIIRAAVREKGLPIVELL 901
Db 712 HSKRMKIAMGSAALRNLMANRPAYKADANIMSPGSLPSLH--VRKQKALEELDAQ 767
OY 902 -----RIDNDRYVCAVATALRNALDVNRKELIGKYAMRDVLRHLPGGNSNNTASKA 954

```

Db 768 HLETFEDNIDN-----LSPKASHRSQRHKKOSLYGVF-----DTNRHDNRSDN 813  
 QY 955 MSDDVTAVCCTLHEVITKNNENAKALRDAGICEKLVGISKSG-----998  
 Db 814 FNTGMMTVLSPYLTNTVLPSSSSSRGSLDSRSRSLERERIGICGNYPATENPGTS 873  
 QY 999 DKHSFKVYKASQVILNSMWOYRDLRSLYK---KDGWSQYHFVASSSTIERDROBPYSSSR 1055  
 Db 874 SKRGQISTTAQJAKVAKVEYSALHTSOEDRSSGSTTELHCV-----TDERALRRSSAAH 929  
 QY 1056 TPS--ISVRYVSPNNRSASAPASPREMISLEKERTDYECTGSNATYHGANGERTSRDAM 1113  
 Db 930 THSNYNTKTSKNSNRKCSMPYAKLE--YKRSSNDSILNSVSSSDGYKRGQ--MKPSI 983  
 QY 1114 TAONTGISTLYRNSGAPAEIDIKHNOVSAOPVPOEPSKQDYETOPONSTRANDE 1169  
 Db 984 ESYSEDESKF-CSTGYPPADLAKIHSAHMD-----NGELDTPTNYSIKYDE 1034

RESULT 12  
 US-07-741-940-2  
 : Sequence 2, Application US/07741940  
 : Patent No. 5352775  
 : GENERAL INFORMATION:  
 : APPLICANT: ALBERTSEN, HANS  
 : APPLICANT: ANAND, RAKESH  
 : APPLICANT: CARLSON, MARY  
 : APPLICANT: GRODEN, JOANNA  
 : APPLICANT: HEDGE, PHILIP J.  
 : APPLICANT: JOSIYN, GEOFF  
 : APPLICANT: KINZLER, KENNETH  
 : APPLICANT: MARKHAM, ALEXANDER F.  
 : APPLICANT: NAKAMURA, YUSUKE  
 : APPLICANT: THLIVERIS, ANDREW  
 : TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC  
 : TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS  
 : NUMBER OF SEQUENCES: 94  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Banner, Birch, McKie & Beckett  
 : STREET: 1001 G Street, NW  
 : CITY: Washington  
 : STATE: D.C.  
 : COUNTRY: USA  
 : ZIP: 20001-4598  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : COMPUTER: IBM PC compatible  
 : OPERATING SYSTEM: PC-DOS/MS-DOS  
 : SOFTWARE: Patent Release #1.0, Version #1.25  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/07/741,940  
 : FILING DATE: 19920109  
 : CLASSIFICATION: 435  
 : ATTORNEY/AGENT INFORMATION:  
 : NAME: Kagan, Sarah A.  
 : REGISTRATION NUMBER: 32,141  
 : REFERENCE/DOCKET NUMBER: 1107.035574  
 : TELECOMMUNICATION INFORMATION:  
 : TELEPHONE: 202-508-9100  
 : TELEFAX: 202-508-9299  
 : INFORMATION FOR SEQ. ID NO: 2:  
 : SEQUENCE CHARACTERISTICS:  
 : LENGTH: 2843 amino acids  
 : TYPE: AMINO ACID  
 : TOPOLOGY: linear  
 : MOLECULE TYPE: protein  
 : US-07-741-940-2

Query Match 3.0%; Score 192; DB 1; Length 2843;  
 Best Local Similarity 18.5%; Pred. No. 7.8e-05;  
 Matches 232; Conservative 163; Mismatches 429; Indels 432; Gaps 49;

QY 23 SASKTSLSL-----GLNLSNGGSETE--TTSAILASVKEQF-----LQ 61  
 Db 102 SVSSNRSGECSVPVPMGSPRRGFVNGSRESTGYLLELEKERSLLDLDEKEEKDXYAQ 161  
 QY 62 FEELITRELEA-----ERQIVASQLE-----RCKLGETSGSMSSSAEFOFQMS 106  
 Db 162 LQNLTKRISLPLTENFSLQTLTRQLEYENARQIVAMEEQLGTCOMERAKQRIARI 221  
 QY 107 QDGKDIIEDELTTGLVDSCIRSLQESGILDPDQYSTEGERPSLLSOGALDINSKEGSEF 166  
 Db 222 QOIEKDI-----LRITROLQS-----QATEERSS-----QNKHETGSH 255  
 QY 167 QVPASHVNOTLAGEPTTPSQLPARGTQARATQSGFSGTSSRACHLAGPEAPPPPPPP 226  
 Db 256 D--AEQNEGGQVGEIN-----MATSGN--GGGSTRMDH-----286  
 QY 227 REPFAPSLGSAEHLDPAPAAAAALYYSSSTLPAPRPGSPFLAPQGGSPFKLQRGSA 286  
 Db 287 -----ETASVLSSSTSHAPRRLTSHLG-----TKVEMVYS- 317  
 QY 287 PEGATYAAPRGSSPKQSPRLAKYSTSPINIVVSSAGLSPIRITSPPTVQSTISSPSI 346  
 Db 318 -----LLSMLGTHDKDMSRTLLAMSSQDSCTISMRQSGCLPLI-----357  
 QY 347 HOLSTIGTYATLSPTKRLVHASEQYKHSQELVATATLQRRGSLAAGSRASYSQHG-- 404  
 Db 358 -----QLHGNKDS-----VLGNSGSKAEARARS--AAAHNIT 391  
 QY 405 HLGPE-----LRALQSEHHIDPIYEDRYQKPPMSLSQSGDPLPPAHGTYYR 454  
 Db 392 HSQPPDKRGRREIRVLLHLEQIRAYCETCWEQEAHEBGM-----DQDKNMP-----439  
 QY 455 TSTAASSPGVDVSPLORTGQSGHCPQMAAATFQRASVAGASNADRYRLOQCPSEVS 514  
 Db 440 -----APVEH-----QICPAY-- 450  
 QY 515 PYKSGPALPREGTLARSPSIDSIQKDP--REFGWRDPELPEYIQLQ-----HQ 562  
 Db 451 -----CYLMKLSFDEHRAHMANELG-----GLQAIAMELLOYDCMYGLTNDHY 493  
 QY 563 FFSVQSNAAATYLOHLCFQDNKKAEI--RRQGIQLVLDLHRTYVRSAGLARILYV 621  
 Db 494 SITLRRYGMALTNTLTFEDVANAKATLCMKGCMRALVLAQLSESEDLQOYVASVLRNLSW 553  
 QY 622 GRANDNKTALKNCGGIPALVRLKTTDLRELVTGVLNLS--CDALMPT--TQDA 678  
 Db 554 -RADVNSKKTIREVGSVAIMECALFEVKESTLSVLSALNNLHACHTENKADICAVDGA 612  
 QY 679 IAVLTNAVLIIPHSQWENSPLODRKIQLHSSQVLRNATGCLRNYS--AGEEARRRMR 735  
 Db 613 IAFIVGTILY-----RSQNTLATIESGGGILRNYSILATNEDHRQLIRE 658  
 QY 736 CDGLDALLIYQSLGSSSEIDSKTYENCVCILRLSLRLAELISOGHMGTDDELGLIC 795  
 Db 659 NNCLOTLLQHLKSHSL-----TIVSNACGTIWNLSR-----690  
 QY 796 GBANCKDAESSGCMGKKKKKKKSDOMDGVGPLDPCAPRKGITOMLHPSTIVKYITLLS 855  
 Db 691 ---NPKDEA-----LMDMGAVS--MLNLI 711  
 QY 856 ECSNPDTEGAAGALONIAA-----GSMKWSYIIRAAVKEKGLPIVELL 901  
 Db 712 HSKHKMIAMGSAALIRNMANRPAKKYKADANTMSPGSLH---VRKQALEAELEDA 767  
 QY 902 -----RIIDDRVYCAVATALRMALDVRKKEILGKYAMMDLVHRLPGGNSNNTSKA 954  
 Db 768 HLETFEDNIDN-----LSPKASHRSQRHKKOSLYGVF-----DTNRHDNRSDN 813  
 QY 955 MSDDVTAVCCTLHEVITKNNENAKALRDAGICEKLVGISKSG-----998  
 Db 814 FNTGMMTVLSPYLTNTVLPSSSSSRGSLDSRSRSLERERIGICGNYPATENPGTS 873  
 QY 999 DKHSFKVYKASQVILNSMWOYRDLRSLYK---KDGWSQYHFVASSSTIERDROBPYSSSR 1055



Db 874 SKRGLOISTTAQIAKVMEEVSAIHTSQEDRSSGTTTELHCY---TDERNALRRSSAAH 929  
 QY 1056 TPS--ISPVRSVNNNSASAPSPREMIISLKERKTDYECTGSNATYHGAKEHTSKRDM 1113  
 Db 930 THSNNTYNTFTKSENSNMTCSMPYAKLE---YKRSSNDLSVSSNDGCGKRGQ---MKPSI 983  
 QY 1114 TAONTGISTLYRNSYGAPAEIDIKHNOVSAQVPQEPSPRKYDETQFONSTENYDE 1169  
 Db 984 ESVSEDESEKRF-CSTGYQPADLAHKIHSANHMDD---NDGELDPINISLAKYSDE 1034  
 RESULT 13  
 US-08-289-548A-2  
 ; Sequence 2, Application US/08289548A  
 ; Patent No. 5648212  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ALBERTSEN, HANS  
 ; APPLICANT: ANAND, RAKESH  
 ; APPLICANT: CARLSON, MARY  
 ; APPLICANT: GRODEN, JOANNA  
 ; APPLICANT: HEDGE, PHILIP J.  
 ; APPLICANT: JOSLYN, GEOFF  
 ; APPLICANT: KINZLER, KENNETH  
 ; APPLICANT: MARKHAM, ALEXANDER F.  
 ; APPLICANT: NAKAMURA, YOSUKE  
 ; APPLICANT: THLIVERIS, ANDREW  
 ; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC  
 ; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS  
 ; NUMBER OF SEQUENCES: 102  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Banner & Allegretti, LTD  
 ; STREET: 1001 G Street, NW  
 ; CITY: Washington  
 ; STATE: D.C.  
 ; COUNTRY: USA  
 ; ZIP: 20001-4598  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: Patentin Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/289,548A  
 ; FILING DATE: 12-AUG-1994  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Kagan, Sarah A.  
 ; REGISTRATION NUMBER: 32,141  
 ; REFERENCE/DOCKET NUMBER: 1107,46943  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 202-508-9100  
 ; TELEFAX: 202-508-9299  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 2843 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-289-548A-2  
 Query Match 3.0%; Score 192; DB 1; Length 2843;  
 Best local Similarity 18.5%; Pred. No. 7.8e-05;  
 Matches 232; Conservative 163; Mismatches 429; Indels 432; Gaps 49;

QY 107 QDGGKIDIEDLTGELVDSICRSLQSGIILPODYSTGERPSILSOLQNSKPEGSF 166  
 Db 222 QOIERDI-----LIRIOLLQS-----QATEAERS-----QNNHETGSH 255  
 QY 167 QYPASYSNQTALAGETTPSOLPARGIOARATGOSFSGQITTSRAGHLAGEPPAPPPPP 226  
 Db 256 D---AERONEOGVGEIN-----MATSQN--GQGSTTRMDH----- 286  
 QY 227 REPAPLSGSAFHLPDAPPAALAAALYSSSTLPAPPGSGPLAAGSGPTKLORGSA 286  
 Db 287 -----ETASVLSSTSHAPRLTSHLG-----TKVEMYSS- 317  
 QY 287 PEGATYAAPRGSSPKQSPSRLAKSYSTISPPINIVSSAGSLPIRSTPYVQSTISSPT 346  
 Db 318 -----LISMLGTHDKDDSRFTLLAMSSSDSCISMRSGCLPLLI----- 357  
 QY 347 HOLSTITVATLSPTKRLVHASQYSKHSQELATATIQRPGLSAGRSYSSOHG-- 404  
 Db 358 -----OLHGNDDKS-----VLGNSRGSKEPARARAS--NALNITI 391  
 QY 405 HLGPE-----LRALQSPREHIDPIEDRYOKRPMKSLQSOGDPLPRAHTGTYR 454  
 Db 392 HSQPDCKRGREIRVHLHLEQIRAVCETCWEOMEAHEBGM---DQKKNPW----- 439  
 QY 455 TSTAPSSPGVDSVPLQRTGSOHGPNMAAATFORASTAAGPASYADPYROLQYCPSEVS 514  
 Db 440 -----APVEH-----QICPAV-- 450  
 QY 515 PYSKSGPALPREGTLARSPSIDSIQKDE-REFGWRDPLPEVIOMLQ-----HQ 562  
 Db 451 -----CYLMLSTDEEHRHAMNELG---GLQAIALLQVDCMYGLTNDHY 493  
 QY 563 PPSVOSNAAYLQHLFCFQDNKIKAEI--RRQGIOLVLDLDMHTEVHRSAGCALRLVLY 621  
 Db 494 STLRRYAGMLTNLTFQGVANKATLCGCMKCRALVQLKSESEDLQOIVASYALRLNSW 553  
 QY 622 GKANDNKIALKNGGITAPLRLKTTDLRELYLVGLWNLS-CDALKMPI--IQDA 678  
 Db 554 -RADVNSKKTLEWGSVVALMECALVEKRESTLSVLSALNLACHTENADICAADGA 612  
 QY 679 LAVLNAVYIIPHSWENSPLODDRKIOLHSSOVLMNATGCLRNYS--AGEEARRRRE 735  
 Db 613 LAFVLGTLTY-----RSQNTLAIIESGGIILRNYSLSLATIEDHROLIRE 658  
 QY 736 CDGLTDALLYIQSALGSSEIDSKTENCVCILNLSYRLAETSGQGHMGTDELGLLC 795  
 Db 659 NNCLQTLQHLKHSLSL-----TIVNACGTIMNLSAR----- 690  
 QY 796 GEANGKDAESSGCKKKKKKKKSQDQWGVGPLPCAPPRPGIQMLHPSLYKTYLTLS 855  
 Db 691 ---NPKDQEA-----LMDGAVS--MLKNLI 711  
 QY 856 ECSNPDTLEGAGALONLAA-----GSMKSVYIRAAVKEKGLPIVLELL 901  
 Db 712 HSKHKIMAGSAAALRLNLMANRPAYKDNIMSPSSLSPLH---YAKKALALELDAQ 767  
 QY 902 -----RIDNDRVCAVATALRNALDVRNKELGKAYAMDVLHRLPGNNSNNTASKA 954  
 Db 768 HLETFEDNIDN-----LSPKASHRSKORHKOSLYGYVF-----DTNRHDNRSDN 813  
 QY 955 MSDDTVTAVCCITLHEVITKKNMNAKALRDAGIEKLVGISKSG----- 998  
 Db 814 FVTGNMTVLSPLYLNTTVLPSSSSSSSGSLDSSRSKEDRSLERGRIGLGNYPATENPCTS 873  
 QY 999 DKHSPPVKAASQVLSNMWQYRDLRLSKYK---KDWMSQYHFAVSSSTLERDRORYSSSR 1055  
 Db 874 SKRGLOISTTAQIAKVMEEVSAIHTSQEDRSSGTTTELHCY---TDERNALRRSSAAH 929  
 QY 1056 TPS--ISPVRSVNNNSASAPSPREMIISLKERKTDYECTGSNATYHGAKEHTSKRDM 1113  
 Db 930 THSNNTYNTFTKSENSNMTCSMPYAKLE---YKRSSNDLSVSSNDGCGKRGQ---MKPSI 983

QY 1114 TAONTGISTLYRNSYGAPAEIDIKHNOVSAOPVPOEPSKDYETQOPONSTRANDE 1169  
DB 984 ESYSEDESKF-CSYGQYPADLAHKIHSANMMD-----NDGELDTPTINYSIKYSDE 1034

RESULT 14  
US-08-452-654-2  
Sequence 2, Application US/08452654

Patent No. 5691454

GENERAL INFORMATION:

APPLICANT: ALBERTSEN, HANS

APPLICANT: ANAND, RAKESH

APPLICANT: CARLSON, MARY

APPLICANT: GRODEN, JOANNA

APPLICANT: HEDGE, PHILIP J.

APPLICANT: JOSIYN, GEOFF

APPLICANT: KINZLER, KENNETH

APPLICANT: MARKHAM, ALEXANDER F.

APPLICANT: NAKAMURA, YUSUKE

APPLICANT: THILVERIS, ANDREW

TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC

TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS

NUMBER OF SEQUENCES: 94

CORRESPONDENCE ADDRESS:

ADDRESSEE: Banner, Birch, McKie & Beckett

STREET: 1001 G Street, NW

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20001-4598

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/452,654

FILING DATE: 25-MAY-1995

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/741,940

FILING DATE: 08-AUG-1991

ATTORNEY/AGENT INFORMATION:

NAME: Kagan, Sarah A.

REGISTRATION NUMBER: 32,141

REFERENCE/DOCKET NUMBER: 1107.035574

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-508-9100

TELEFAX: 202-508-9299

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 2843 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-452-654-2

Query Match 3.0%; Score 192; DB 1; Length 2843;

Best Local Similarity 18.5%; Pred. No. 7.8e-05;

Matches 222; Conservative 163; Mismatches 429; Indels 432; Gaps 49;

QY 23 SASERTSLSP-----GLNTSNGDSETE-TTSALIASVKEOE-----LQ 61

DB 102 SVSRSGECSPVPMGSPRRGFGVNGRESSTGYDELEKERSLLADLDKREKKDMYVQ 161

QY 62 FERLTRELEA-----ERQIVASOLF-----RCKIGSETGSMSSMSAEQFOWOS 106

DB 162 LQNTLRKIDSLPTENSLQTDLTFRQLEYEARQIRVAMEEQGLTCQDMEKRAQRIARI 221

QY 107 ODGOKIDELELTGLBELVDCISLQESGLDPDQDSTGRBRLSQSALQLSKREGSF 166

DB 222 QQTEKDI-----LRIRQLQS-----QATEAERSS-----QNKHETGSH 255

QY 167 QYASYSNQTALGETTTSQLPARGCTARATGQSFSGTTSRACHLAGPEPAPPPPP 226

DB 256 D--AERONEGQGVGEIN-----MATSGN-GQSTTRMDH----- 286

QY 227 REPFASLSAFHLPAPPAAMAAALYYSSSTLPAPPGGSPPLAPGGSPKTLQRGSA 286

DB 287 -----ETASVLSSTSHSPRRLTSHLG-----TKVMYVS- 317

QY 287 PEGATYAAPRGSSPKQSPSLAKSYSTSSPINIYVSAGLSPIRYTSPTVQSTISSPI 346

DB 318 -----LLSMGLTHDKDDMSRTLLAMSSSODSCISMRQSGCLPLI----- 357

QY 347 HQLSSTIGTVALSPKRLVHASPOYSKHSQBLVATLQRPGLSIAAGSRASYSOQG- 404

DB 358 -----QILHGNDRKS-----VLLGNSRGSKEARARAS-ALAHNII 391

QY 405 HLGPE-----LRAQSPENHIDPIYEDRYQKPPMSLSQSGDPLPAHTGYR 454

DB 392 HQPPDKRRRRLRVLHLEQIRATCETCWEQEAHERGM-----DQKNPMP----- 439

QY 455 TSTAPSSPGVDSVPLQRTGSGHQGNAAAATFFQASVYAGPASNADYDROLQYCPYES 514

DB 440 -----APVEH-----QICPAV-- 450

QY 515 PYSKSGPALPPEGTLARSPIDSIOKDP-RENGWRDPELPYIQLQ-----HQ 562

DB 451 -----CYLMKLSFDEEHRHAMELNG-----GLQALAEILQYDCEMYGLTNDHY 493

QY 563 PPSVQNAAYLQHLQCFQDNKIKAEI-RRQGGIQLLVLDLHRMEVRSAGALRNLY 621

DB 494 SITLRVYAGMALTNLTFQDVANKATLCMSKGMRAVLQAKSESDDLQOVLASVLRNLSW 553

QY 622 GRANDNKIALKNGGIPALVYLLKRTTDLTRELVTGVLNLS-CDALMPTI--IQDA 678

DB 554 -RADVNSKKTREVSVAIMECALVEKKESTKSVLSALNLSAHCENKADICAVGDA 612

QY 679 LAVLTNAVYIIPHSWENSLQDRKIOIHSQVLENAGCIRANS--AAEARRRRRE 735

DB 613 LAFELVGTLY-----RSQNTLALIESGGGILRNSSSLATVEDRQILRE 658

QY 736 CDGLTDALLYIQSALGSSEIDSKTVENCVCILRNLSYLAETSOQOHMTDELQILC 795

DB 659 NNCLOTLLQHLKSHSL-----TIVSNACGTLMNLSR----- 690

QY 796 GRANGKDAESSCGWKKKKKKKSQDQMDGVGLPDCAPPGIOMLWHPSTVYKYLILLS 855

DB 691 ---NPKDQEA-----LWDMGAVS-MKXNLI 711

QY 856 ECSNPDTEGAAGALONTAA-----GSMKWSYIIRAAYKREGLPIVLELL 901

DB 712 HSKHKIMAGSAAALRNLMANRPAYKRDANIMSPGSSLSH-----YRKQALAELEDAQ 767

QY 902 -----RIDNDRVYCAVATALRNALDVRNKEILGKYAMDOLVHRLPGGNSNNTASKA 954

DB 768 HLSETFDIND-----LSPKASHRSKORHKOSLYGDVYF-----DTNRHDDNSDN 813

QY 955 MSDDTVTAVACCTIHEVITRKNNENAKALRDAGICTEKLIGISKSG----- 998

DB 814 FNTGNMTVLSPLYLNTTVLPSSSSSSSGSLDSSRSSEKDRLEREIRIGLGNYPATENEGTS 873

QY 999 DKHSPPVYKKAASOVYLSNMOYRDLRSLYK--KDGWSQYHRYVASSSTLERROPYSSR 1055

DB 874 SKRGQITSTAQAQIAKVHEVSAIHTSGEDSSSGSTTELHCY-----IDERNALRRSSAAH 929

QY 1056 TPS--ISPVVSPNNRSASAPAPREMIISLKERKTYDECTGSNATYTGANGCHTSRDM 1113

DB 930 THSNNTYNTKSENSNRTCSMPYAKLE--YKSSNDLSINSSNDYGGKRGQ--MKPSI 983

QY 1114 TAONTGISTLYRNSYGAPAEIDIKHNOVSAOPVPOEPSKDYETQOPONSTRANDE 1169

DB 984 ESYSEDESKF-CSYGQYPADLAHKIHSANMMD-----NDGELDTPTINYSIKYSDE 1034

## RESULT 15

US-08-370-235A-2

Sequence 2, Application US/08370235A

Patent No. 5910418

GENERAL INFORMATION:

APPLICANT: VOGELSTEIN, BERT

APPLICANT: KINZLER, KENNETH W.

APPLICANT: HILL, DAVID E.

APPLICANT: JOHNSON, KAREN A.

TITLE OF INVENTION: ANTIBODIES AND ASSAYS FOR DETERMINING

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: BANNER &amp; MITCHELL, LTD.

STREET: 1001 G STREET, N.W.

CITY: WASHINGTON

STATE: DC

COUNTRY: US

ZIP: 20001

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/370,235A

FILING DATE: 01-JAN-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: KAGAN, SARAH A.

REGISTRATION NUMBER: 32,141

REFERENCE/DOCKET NUMBER: 01107.48688

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202 508 9100

TELEFAX: 202 508 9299

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 2843 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-370-235A-2

Query Match 3.0%; Score 192; DB 2; Length 2843;

Best Local Similarity 18.5%; Pred. No. 7.8e-05;

Matches 232; Conservative 163; Mismatches 429; Indels 432; Gaps 49;

QY 23 SASKTSLSLP-----GLNTSNGDSETE-TTSAILASYKEOE-----LQ 61

Db 102 SVSRSGECPFPWGFPRGFVNGSRESSTGYLELEKERSLLADLDKREKEDWYAO 161

QY 62 FERITRELEA-----ERQIVASOLE-----RCKLGSEMSMSMAEPOFOWOS 106

Db 162 LQNTKRIDLPLETFENSLQTDLTIRLEFEARQIRVAMEEOLGTQDMEKRAQRRIARI 221

QY 107 QDQOKDIEDLTGLLELVDSICRSLQESGLDPQDSTGERPRLLOSAQLNKEGSGF 166

Db 222 QQLEKDI-----LRIRQLLOS-----QATEERSS-----QNKHETGSH 255

QY 167 QYPAVSHNQTLAIGETTPSLPARSTQARATGQSFQGTSTPRAGHAGPEPAPPPPP 226

Db 256 D-----AERONGGQGVGEIN-----MATSGN-GQGSSTTMDH----- 286

QY 227 REPPAFLSGAFLHPAPPAALALYSSSTLPAPPPGSPPLADPGSGPTKLQRGSGA 286

Db 287 -----ETASVSSSTSHAPRRLTSHLG-----TKVENWYS- 317

QY 287 PEGATYAAPRGSSPKOSPSHLAKSYSTSSPINIVVSSAGLSPIRTVTSPTVOSTISSSPI 346

Db 318 -----LLSMGLGTHDKDDMSKTLIAMSSQDSCISMROSGCLPLLI----- 357

QY 347 HOLSTIGTATLSPTRKIVYHASEOYSKHSQELVATATLQRPISLAAGRSASYSQH-- 404

Db 358 -----QLHGNDDKDS-----VILGNSRSGSKARARAS--AALHNII 391

QY 405 HLGE-----LRALQSPHHIDPIYEDRVYOKPPMRSLQSGPLPAHGTFR 454

Db 392 HSQDDKRGRRREIRVHLHLLEQIRAYCETCWEQEAHPCH-----DDKPNP----- 439

QY 455 TSTAASSGVDSVPLQRTSGOHGPQNAATFORASVYAGPASNVDAPYROLQYCPVSES 514

Db 440 -----APVEH-----QICPAV----- 450

QY 515 PYSKGPALPPESTLARSPIIDSIQDP--REFGWRPELPEYIOMQ-----HQ 562

Db 451 -----CVLMKLSFDEEHRHANEIG-----GLQALAEILLQVDCENYGLTNDHY 493

QY 563 FPSVOSNAAAYLOHLCFGDNKIKAEI--RROGGIOLVLDLDRHMEVHRSAGALRNLY 621

Db 494 SILRRYAGMALNLTFFGDVANKATLCSMKGCMRALVLAQLKSESDLOQVYASVLRNLSW 553

QY 622 GKANDNKIALKNGGIPALVRLRKTDTLEIRELVYVLMNLS--CDALKMPI--IDA 678

Db 554 -RADVNSKKTREVGSVYKALMECALEVKEESTLKVSLAMNLSAHTENKADICAVDGA 612

QY 679 LAVLTNAVLIIPHSQWENSPLODDRKTOLHSSQVLRNATGCLNNVS--AGEARRRRE 735

Db 613 LAFVGLTY-----RSQVTLAIIESGGILAEVSLATATNDDHQIARE 658

QY 736 CDGLTDLALVYIOSALGSSEIDSKTYENCJILRNLSYLAETSOQGMHDELDGLLC 795

Db 659 NNCLOTQLQHLKSHSL-----TIVSNACGLTMNLSAR----- 690

QY 796 GEANGDAASSGCGWKKKKKKKSDQDGVPLPCDAEPKGIOMLMHPSIVKPYLTLS 855

Db 691 ---NPKDOEA-----LMDGAVS--MLKNLI 711

QY 856 ECSNPDTLEGAGALONLAA-----GSMKSVYIRAAVREKGLPLVELL 901

Db 712 HSKHKMIAGSAAALNRLNANRPAYKKDANINSPSSSLPSLH--VKRQALELEDAQ 767

QY 902 -----RIDNDRVYCAVATALRNALDVRNRELIGKYAMRDVLRPLGNGNSNTASKA 954

Db 768 HLETFDNDN-----LSPKASHRSKQHRKQSLYGVYF-----DTNRHDNRSDN 813

QY 955 MSDDTYAVCCCTLLHEVYTKNMENAKALRDAGIEKLVGISKSG----- 998

Db 814 FNTGMATVLSPYLNTTVLPSSSSSRGSLDSRSERKDRSLERERIGLGNYPATENPETS 873

QY 999 DKHSPVYVAAQOVLSNMQYRDLRSYK--KQMSQYHYFASSSTIERDRORYSSR 1055

Db 874 SKRGLOISTTAQOIAKVMEEVSAIHTSOEDRSSGSTTELDHV---TDERNALRRSSAAH 929

QY 1056 TPS--ISPVYVSPNRRSAPASPPEMISLKERKDYECTGSMATYHAKGEHSTRKDM 1113

Db 930 THSNYNYNFTKSPNSNRKTSMPYAKL---YKRSSNDLSNYSNDGVYKRRQ---MKPST 983

QY 1114 TQONTGISTLYNYSYGAPEADIKHNQVSAOPVPOEPRKDYETQYOPONSTRNRYDE 1169

Db 984 ESYSEDDSKF--CSYGQYPADIAHKIHSANHMD---NOGELDLPINYSILKYSDE 1034

Search completed: July 19, 2001, 16:03:58  
Job time: 188 sec

Fri Jul 20 10:09:40 2001

us-09-501-171a-4.ra1

Page 16

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using SW model

Run on: July 19, 2001, 16:03:20 / Search time 14.73 Seconds

(without alignments)  
2848.810 Million cell updates/sec

Title: US-09-501-171A-4

Perfect score: 6377

Sequence: 1 MFARKPPGAAPLGAMPVDDQ.....PYSELNVENTSHYPASPSDW 1225

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 93435 segs, 34255486 residues

93435

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt\_38:\*

# SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	ID	Description
1	1423.5	22.3	969	1 ARVC_MOUSE	P08203 mus musculu
2	1417	22.2	962	1 ARVC_HUMAN	O00192 homo sapien
3	1375.5	21.6	911	1 P120_MOUSE	P30999 mus musculu
4	256.5	4.0	1185	1 DRPL_HUMAN	P54259 homo sapien
5	226.5	3.6	1183	1 DRPL_MOUSE	P54258 rattus norv
6	205	3.2	721	1 YK82_MYCTU	Q10690 mycobacteri
7	202	3.2	1487	1 ICP4_HSVB	P28925 equine herp
8	201	3.2	817	1 VRP1_YEAST	P37370 secccharomyc
9	200.5	3.1	1902	1 SMF1_HUMAN	O14497 homo sapien
10	200	3.1	2845	1 APC_MOUSE	O61315 mus musculu
11	197	3.1	1461	1 IE18_PRTIF	P11675 pseudorabie
12	195	3.1	1140	1 VM96_YEAST	O04893 secccharomyc
13	194.5	3.1	2842	1 APC_RAT	P70478 rattus norv
14	194	3.0	660	1 VHL1_EBV	P03181 epstein-bar
15	194	3.0	1575	1 STJ1_HUMAN	O43426 homo sapien
16	194	3.0	2843	1 APC_HUMAN	P25054 homo sapien
17	193.5	3.0	781	1 CTNB_RAT	O94082 rattus norv
18	191.5	3.0	781	1 CTNB_HUMAN	P35222 homo sapien
19	191.5	3.0	1083	1 T2D3_HUMAN	O00268 homo sapien
20	191.5	3.0	1896	1 RBP1_DROME	P04052 drosophila
21	190.5	3.0	1744	1 TENS_CHICK	O04205 gallus gall
22	190	3.0	1487	1 ICP4_HSVB	P17473 equine herp
23	189.5	3.0	781	1 CTNB_MOUSE	O02248 mus musculu
24	188.5	3.0	1273	1 WEI1_YEAST	P38968 secccharomyc
25	187	2.9	578	1 VAC8_YEAST	P39968 secccharomyc
26	185.5	2.9	2517	1 NCR2_HUMAN	O94618 h nuclear r
27	183.5	2.9	1125	1 SZ4C_HUMAN	P33992 homo sapien
28	183.5	2.9	1306	1 MSB2_YEAST	P32334 secccharomyc
29	183.5	2.9	1429	1 EXPA_DROME	O07436 drosophila
30	183.5	2.9	3828	1 TRX_DROVI	O24742 drosophila
31	183	2.9	781	1 CTNB_XENLA	P26233 xenopus lae
32	182.5	2.9	813	1 ARM_MUSDO	O02451 musca domes
33	182.5	2.9	1970	1 RBP1_HUMAN	P24928 homo sapien

## ALIGNMENTS

34	182.5	2.9	1970	1 RBP1_MOUSE	P08775 mus musculu
35	182	2.9	1075	1 NEC3_HUMAN	O12968 homo sapien
36	181.5	2.8	818	1 CTNB_URECA	P35224 urechis cau
37	180	2.8	1443	1 E75C_DROME	P13055 drosophila
38	180	2.8	1647	1 SN24_HUMAN	P51532 homo sapien
39	179.5	2.8	1122	1 HD45_HUMAN	O94016 homo sapien
40	179.5	2.8	1367	1 AMYH_YEAST	P08640 secccharomyc
41	178.5	2.8	1365	1 SUZ2_DROME	P25172 drosophila
42	178	2.8	2090	1 N214_HUMAN	P35658 homo sapien
43	177.5	2.8	2805	1 MAPA_HUMAN	P78559 homo sapien
44	176.5	2.8	1077	1 HLES_DROME	O02308 drosophila
45	176	2.8	701	1 CG1_HUMAN	O13495 homo sapien

RESULT 1  
ARVC\_MOUSE STANDARD; PRT; 969 AA.

AC P08203: 01-OCT-2000 (Rel. 40, Created)  
DT 01-OCT-2000 (Rel. 40, Last sequence update)  
DT 01-OCT-2000 (Rel. 40, Last annotation update)  
DE ARMADILLO REPEAT PROTEIN DELETED IN VELO-CARDIO-FACIAL SYNDROME  
DE HOMOLOG (FRAGMENT).  
GN ARVCF  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_Taxid=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Starzynski-Powitiz A., Kaufmann U., Urbich C., Zupplinger C.;  
RT "The armadillo repeat protein ARVCF, a candidate for the velo cardio  
RT facial syndrome forms complexes with cadherin."  
RL Submitted (JUN-1999) to the EMBL/GenBank/DBD databases.  
CC -!- FUNCTION: INVOLVED IN PROTEIN-PROTEIN INTERACTIONS AT ADHERENS  
CC JUNCTIONS (BY SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE BETA-CATENIN FAMILY.  
CC -!- SIMILARITY: CONTAINS 10 ARM REPEATS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: AJ23418; -!- NOT ANNOTATED\_CDS.  
DR MGD: MGI:109620; Arvcf.  
DR PROSITE: PS50176; ARM\_REPEAT; 3.  
KW Repeat; Coiled coil.  
FT NON\_TER 1  
FT DOMAIN 1  
FT REPEAT 328  
FT REPEAT 377  
FT REPEAT 419  
FT REPEAT 462  
FT REPEAT 503  
FT REPEAT 521  
FT REPEAT 570  
FT REPEAT 616  
FT REPEAT 634  
FT REPEAT 674  
FT REPEAT 720  
FT REPEAT 721  
FT REPEAT 725  
FT REPEAT 770  
FT REPEAT 814  
FT REPEAT 105378 MW; 91130069E484C5B5 CRC64;  
SQ

COILED COIL (POTENTIAL).  
ARM 1.  
ARM 2.  
ARM 3.  
ARM 4.  
ARM 5.  
ARM 6.  
ARM 7.  
ARM 8.  
ARM 9.  
ARM 10.  
ARM 11.  
ARM 12.  
ARM 13.  
ARM 14.  
ARM 15.  
ARM 16.  
ARM 17.  
ARM 18.  
ARM 19.  
ARM 20.  
ARM 21.  
ARM 22.  
ARM 23.  
ARM 24.  
ARM 25.  
ARM 26.  
ARM 27.  
ARM 28.  
ARM 29.  
ARM 30.  
ARM 31.  
ARM 32.  
ARM 33.  
ARM 34.  
ARM 35.  
ARM 36.  
ARM 37.  
ARM 38.  
ARM 39.  
ARM 40.  
ARM 41.  
ARM 42.  
ARM 43.  
ARM 44.  
ARM 45.  
ARM 46.  
ARM 47.  
ARM 48.  
ARM 49.  
ARM 50.  
ARM 51.  
ARM 52.  
ARM 53.  
ARM 54.  
ARM 55.  
ARM 56.  
ARM 57.  
ARM 58.  
ARM 59.  
ARM 60.  
ARM 61.  
ARM 62.  
ARM 63.  
ARM 64.  
ARM 65.  
ARM 66.  
ARM 67.  
ARM 68.  
ARM 69.  
ARM 70.  
ARM 71.  
ARM 72.  
ARM 73.  
ARM 74.  
ARM 75.  
ARM 76.  
ARM 77.  
ARM 78.  
ARM 79.  
ARM 80.  
ARM 81.  
ARM 82.  
ARM 83.  
ARM 84.  
ARM 85.  
ARM 86.  
ARM 87.  
ARM 88.  
ARM 89.  
ARM 90.  
ARM 91.  
ARM 92.  
ARM 93.  
ARM 94.  
ARM 95.  
ARM 96.  
ARM 97.  
ARM 98.  
ARM 99.  
ARM 100.

Query Match 22.3%; Score 1423.5; DB 1; Length 969;  
Best Local Similarity 34.4%; Pred. No. 7.4e-61;  
Matches 391; Conservative 153; Mismatches 332; Indels 261; Gaps 42;

```

OY 47 TTSAILASVKEOEIOLERLITRELEAEROIYASOLERCK-----LGSEFGSMSSMSAEQ 101
DB 1 SAAIILASVKEOEAEFERLITRELEAEROIYASOLERCK-----LGSEFGSMSSMSAEQ 60
OY 102 -FOWSODGOKIDELEITLLELVDSCLNSLOESGLDPDOSTGERPBLLOSALQ 159
DB 61 LVLOGSGSGSASHA-----TMPAPAEVLEET-VTVE 91
OY 160 SKRPSQYRPSVSHNOTLALGETTSPQLPARGTOARATGOSFGQTSNAGH-AGEP 218
DB 92 EDP-GT---PSSHVSIYSEDT-----KRETEKVT-KIYVTTTTRVROYPLGPDG 139
OY 219 AP-----PPPPPREPAPSLGSAFHLDPAP-----PAAAALYSSSTLPA 261
DB 140 LPLDGGP-----LGS---FADGFLDRHYLLRGCGGPAATLSRYHSS----- 180
OY 262 PPRGSPPLAAGSGSPYKQLONGSAPBEGATYAPR-----GSSPQOS---PSRLAKSYT 313
DB 181 -GG---GFPDGPESRDIPSYGLSRGLGVRPRPTGLGPGDGGFTLPGR-REAFPM 233
OY 314 SSPINIVVSSAGLSPIRVTSPTVOSTISSPIHQLSSTIGTYATLSPTKLVASBOYS 373
DB 234 GSE-----SGPPSGRL---PEHQAEPYC---LDDTRSLAADDEGG 270
OY 374 KHSQELVATATLORPGLAAGSRASYSOGHGLGPELRLALQSPHHIDPIYEDRYOKP 433
DB 271 PDLEBDYSTATRRPR-----EYGR-GLARAFEDTADAGELLEER----- 310
OY 434 MRSLSQSGDPLPAHNGTCTRTAPSSPGVSVPLQRTGSOHQPQANAATPORAATA 493
DB 311 -----PP-----PFAATAPLAQ-----PEKSGSLDR----- 333
OY 494 GPASVADPQYOLQYCPSESPYKSGPALPEEGLARSPSIDSIQDPREFGRDDEL 553
DB 334 -----VWRSPVDSTRKEPR---WRDELP 356
OY 554 EVIOMLQHPFSSVOSNAAYLQHLCEGDNKIKAEIRROGGIQLLVLLDRHTEVHSAC 613
DB 357 EYLAIRHPVDPVKANAAYLQHLCEGDNKIKAEIRROGGIQLLVLLDRHTEVHSAC 416
OY 614 GALRLVYKANDNDKIALKNGCGIPALVRLRKTDTLEIRLTVGYLWNSDCLAMP 673
DB 417 GALRLVYKANDNDKIALKNGCGIPALVRLRKTDTLEIRLTVGYLWNSDCLAMP 475
OY 674 IYDQALVLTAAVITPHGWSNPLODDRIQLHSSQYLRNATGCLNRVSSAGEARRM 733
DB 476 IYDQALVLTAAVITPHGWSNPLODDRIQLHSSQYLRNATGCLNRVSSAGEARRM 535
OY 734 RECDGLTALLVYQISALGSSEIDSKTVENCVCILRLSYRLAETSOGOMGTDE--LD 791
DB 536 RECDGLTALLVYQISALGSSEIDSKTVENCVCILRLSYRLAETSOGOMGTDE--LD 791
OY 792 GLTCEANGKDAESSGCKKKKKSSODQMD---GVGRLPDCAEPKGIOMLHPSIVK 848
DB 536 GLTCEANGKDAESSGCKKKKKSSODQMD---GVGRLPDCAEPKGIOMLHPSIVK 848
OY 849 PYLITLSECSNPDTLEGAGALQNLAAAGSMKWSYITRAAVKKEGLPIVLELLIDNDY 908
DB 654 PYLITLSECSNPDTLEGAGALQNLAAAGSMKWSYITRAAVKKEGLPIVLELLIDNDY 713
OY 909 VCAVATALRNALDVNRKELIGKIAMRDVLRHLPBGNNNSNTA--SKAMSDIYAVCCT 966
DB 714 VCAVATALRNALDVNRKELIGKIAMRDVLRHLPBGNNNSNTA--SKAMSDIYAVCCT 770
OY 967 LHEVITTKMNAKALRDAGGIEKLVGSKSGDKHSPVVAASOVYNSMGOYRDLISLY 1026
DB 771 LHEVITTKMNAKALRDAGGIEKLVGSKSGDKHSPVVAASOVYNSMGOYRDLISLY 827
OY 1027 KKDGSQYHFAVSSSTIRDRQRPYSSSRTPSISPV--RVSPNNRSASAPASPRE----- 1079
DB 828 KKDGSQYHFAVSSSTIRDRQRPYSSSRTPSISPV--RVSPNNRSASAPASPRE----- 866
OY 1080 ---MISLKERKTYECTGSNATYHGAKGEHTSRKDAMTAONTGISTLYRNSYGAPE 1133

```

```

DB 887 GYATVDRRERRT-----LQSDST-----GD-TSEKELIRYQGV-----YCGPLE 926
RESULT 2
AC ARVC_HUMAN STANDARD: PRT; 962 AA.
ID 000192;
DT 01-OCT-2000 (Rel. 40, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ARADILLO REPEAT PROTEIN DELETED IN VELO-CARDIO-FACIAL SYNDROME.
GN ARVCF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
NCBI_TaxID=9606;
[1]
RX MEDLINE=97271559; PubMed=9126485;
RA Sirocki H., O'Donnell H., Dasgupta R., Halford S., St Jore B.,
RA Puech A., Parimoo S., Morrow B., Skoultschi A., Weissman S.,
RA Scambler P., Kucherlapati R.;
RT "Identification of a new human catenin gene family member (ARVCF) from
RT the region deleted in velo-cardio-facial syndrome."
RL Genomics 41:75-83(1997).
CC -!- FUNCTION: INVOLVED IN PROTEIN-PROTEIN INTERACTIONS AT ADHERENS
CC JUNCTIONS.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A SHORT FORM AND A LONG FORM
CC (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: FOUND IN ALL THE EXAMINED TISSUES INCLUDING
CC HEART BRAIN, LIVER AND KIDNEY. FOUND AT LOW LEVEL IN LUNG.
CC -!- DISEASE: GENE DELETED IN VELO-CARDIO-FACIAL SYNDROME (VCFS) AND IS
CC HEMIZYGOUS IN ALL VCFS PATIENTS WITH INTERSTITIAL DELETIONS. THIS
CC HEMIZYGOUSITY MAY PLAY A ROLE IN THE ETIOLOGY OF SOME OF THE
CC PHENOTYPES ASSOCIATED WITH VCFS CHARACTERIZED BY A WIDE SPECTRUM
CC PHENOTYPES, INCLUDING CONOTRUNCAL HEART DEFECTS, CLEFT PALATE AND
CC FACIAL DYSMORPHOLOGY.
CC -!- SIMILARITY: BELONGS TO THE BETA-CATENIN FAMILY.
CC -!- SIMILARITY: CONTAINS 10 ARM REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DB EMBL; U51269; AAC51202.1; -.
DB HSSP; Q02248; 2BC7.
DB MIM; 602269; -.
DB MIM; 192430; -.
DB InterPro: IPR000225; -.
DB Pfam: PF00514; Armadillo-seg; 4.
DB ProSite: PS0176; ARM_REPEAT; 3.
KW Repeat; Coiled coil; Alternative splicing.
FT DOMAIN 8 46 COILED COIL (POTENTIAL).
FT DOMAIN 607 623 NUCLEAR LOCALIZATION (POTENTIAL).
FT DOMAIN 608 611 POLY-ARG.
FT REPEAT 348 387 ARM 1.
FT REPEAT 390 429 ARM 2.
FT REPEAT 433 473 ARM 3.
FT REPEAT 468 508 ARM 4.
FT REPEAT 526 565 ARM 5.
FT REPEAT 575 622 ARM 6.
FT REPEAT 646 682 ARM 7.
FT REPEAT 699 738 ARM 8.
FT REPEAT 739 781 ARM 9.
FT REPEAT 782 826 ARM 10.
FT VARSPPLIC 1 69 MEDCUNHSAASILASVKEOEAEFERLITRELEAEROIYASOLERCK-----YCGPLE
FT FT BRAODPGVWGGSGVGLPPLAMQDLTVL -> MPAELR
FT FT (IN SHORT ISOFORM).

```



```

Db 62 RRHONGRFVGDADLERQKFSKDLK-----NG-----PODHNHLLSTPIPMOEPGOL 108
OY 152 SOSALQJNSKREGEFQYPASTYHSNQTALGETTIPSOLPANGTOAR--ATQOSQSGTSTR 209
Db 109 VEIYTE--EDEBEGA-----MSVSVETD-----GTRRFETVYKVKVKTWTR 151
OY 210 AGHLAGEPAPPPPPPREPPAPSLGSAFHLDPAPAPAAAALYYSSSTLPAPRGGSPL 269
Db 152 T-----VQPPV-----MG-----PDLPPVDAASVSNITQTL----- 178
OY 270 AAPGSGPTKLORGSAPEGATYAPRGSSPKOSPSRLAASYSSTSPINIVSAGISPI 329
Db 179 -----GRBFRKNGNGPGP----- 192
OY 330 RVTSPPVQOSTISSPHQJLSSTIGTATLSPTKRLVHASEQYKSHQELXATATLORPG 389
Db 193 -----YVQAGTATLPRNFHYPPDGGRHEDGY-----PG 223
OY 390 SLAAGSRASYSOQHGLCP-----ELRALQSPENHIDPIYEDRVYQKPPMSLSOSOGDPL 445
Db 224 G-----SDNYGSLRVTIRIERTRPSMEGRAPSRD--YGGPPQYRVGSSVD-L 272
OY 446 PPAHTGYRTSTAASSPQVDSV-----PLQRTGSSQHPONAAAATFORASYAAGA 496
Db 273 HRFPPEPYGLDDQSRMKGDDLDYGMMSDYGTARTGTP----- 311
OY 497 SNVADPTYOLOQCPVSPEYKSGPALPPE-----GTLARSPSIDQKD-PR 543
Db 312 -----SDPRRLR-----SYEDMIEEVPDDOYWPAPLAQHERGSLA--SIDSIRKMP 359
OY 544 EFGWDEPLPEVYIOMLOHOPFVSOSNAAYLOHLFCFSDNKKIKAIRQOGTOLVLDLDH 603
Db 360 PSNNRQPEPLPEVIMLGRDLDAVSNAAAYLOHLCYNDKDKTVDATKGPILVGLD 419
OY 604 RMTEVHNSACALNLVYCKANDNKATLAKCGGIPALVRLKRTTDELTRELVTVGLVN 663
Db 420 PKREVHLGACGALKNISFGR--DQDKTAKKNCDEVPALVRLKRAKRAMDLEVTGTLNM 478
OY 664 LSSCDALKMPITDIALVLANVLIIPHSQWENSPLODRKLTOLSSQVLRNATGCLRNVS 723
Db 479 LSSHDSIMEIYDALHALTDEVLIIPHSQWENSPLODRKLTOLSSQVLRNATGCLRNVS 538
OY 724 SAGEBARRRMECDGLDALLVYIQSALGSSEIDSKTYENCVCILRLSYRLAETSQGO 783
Db 539 SERSEARRKRECDGLVALLFIYQAEIGQKDSKTYENCVCILRLSYQVHREIQO-- 596
OY 784 HMGTDLDGLGCEANGDASSGCGWKK--KKKKKSDQDDVGPPLDCAEPKKGLOML 841
Db 597 ---ARVGEALPYVANSSTGPHRASCFCGAKKGGKPPTEDPANDTVDPKRTSPARKIELL 653
OY 842 WHPSIVKPYLTLSSECSNDPTELEGAAGALQMLAAGSMWYSYIRAAVKEKEGLPIVELL 901
Db 654 FOPEVVRITISLTKESNPALLEASAGAIQNLGGRMYGYRISALRQEKALABARELL 713
OY 902 RIDNDRVCAVATALRNALDVRNKKELIGKYAMRDLYHRLPGNNSNNTASKAMSDOTVT 961
Db 714 TSEHERVYKASGALRNLAVDARNKELIGKHAHPNLYKNLPGGQON--SSWNSSEDTVV 770
OY 962 AVCCCLHEVITKMMENKALDAGIEKLVGKSGKDKHSPKTVYKASQVLSNMOYOR 1021
Db 771 SLTINTEVIENLEAKAKLETOGIEKLVLINKS--GNRSEKEVRAALVLOITNGEYKE 828
OY 1022 LRSLYKKGWQSHFVASSSTIERDR--ORPYSSSRTPSISFVSPNNRS 1070
Db 829 LRKPLEKGGWKSDFOVNLNNAASRQSSHSYDSTLPLIDRNQSDNNYS 878

```

```

DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE ATROPHIN-1 (DENTATORUBRAL-PALLIDOLUYSIAN ATROPHY PROTEIN).
GN DRPLA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid:9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cerebellum, and Brain;
RX MEDLINE=96262314; PubMed=8965642;
RA Margolis R.L., Li S.-H., Young W.S., Wagster M.V., Stine O.C.,
RT Kidwai A.S., Ashworth R.G., Ross C.A.;
RT "Drp1a gene (atrophin-1) sequence and mRNA expression in human
RL brain."
RN Brain Res. Mol. Brain Res. 36:219-226(1996).
RP [3]
RP SEQUENCE OF 470-725 FROM N.A.
RC TISSUE=Brain cortex;
RX MEDLINE=93315145; PubMed=8325628;
RA Li S.-H., McInnis M.G., Margolis R.L., Antonarakis S.E., Ross C.A.;
RT "Novel trip1et repeat containing genes in human brain: cloning,
RT expression, and length polymorphisms."
RL Genomics 16:572-579(1993).
CC -1- TISSUE SPECIFICITY: THE LEVELS ARE RELATIVELY HIGH IN THE BRAIN,
CC Ovary, Testis and Prostate. Lower levels are detected in the
CC Liver, Thymus and Leukocytes.
CC -1- POLYMORPHISM: THE POLY-GLN REGION OF DRPLA IS HIGHLY POLYMORPHIC
CC (7 TO 23 REPEATS) IN THE NORMAL POPULATION AND IS EXPANDED TO
CC ABOUT 49-75 REPEATS IN DRPLA PATIENTS. LONGER EXPANSIONS RESULT IN
CC EARLIER ONSET AND MORE SEVERE CLINICAL MANIFESTATIONS OF THE
CC DISEASE.
CC -1- DISEASE: DEFECTS IN DRPLA ARE THE CAUSE OF DENTATORUBRAL-
CC PALLIDOLUYSIAN ATROPHY, AN AUTOSOMAL DOMINANT NEURODEGENERATIVE
CC DISORDER CHARACTERIZED BY A LOSS OF NEURONS IN THE DENTATE
CC NUCLEUS, RUBROM, GLOBUS PALLIDUS AND LYS/BODY. CLINICAL FEATURES
CC ARE MYOCLONUS EPILEPSY, DEMENTIA, AND CEREBELLAR ATAXIA. ONSET OF
CC THE DISEASE OCCURS USUALLY IN THE SECOND DECADE OF LIFE AND DEATH
CC IN THE FOURTH.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch).
CC -----
DR EMBL; D31840; BA00626.1; -
DR EMBL; U23851; AAB50276.1; -
DR EMBL; L10377; -; NOT_ANNOTATED_CDS.
DR HSSP; P00651; 1LRA.
DR MIM; 125370; -
DR InterPro; IPR002951; -
DR PRINTS; PR01222; ATROPHIN.
KW Triplet repeat expansion; Polymorphism.
FT DOMAIN 73 82 SER/GLU-RICH (MIXED CHARGE).
FT DOMAIN 302 305 POLY-PRO.
FT DOMAIN 376 382 POLY-SER.
FT DOMAIN 386 397 POLY-SER.
FT DOMAIN 442 447 POLY-PRO.
FT DOMAIN 479 483 POLY-HIS.
FT DOMAIN 484 497 POLY-GLN.
FT DOMAIN 504 507 POLY-PRO.
FT DOMAIN 564 574 POLY-SER.

```







```

RESULT 7
ICP4_HSVB STANDARD; PRT; 1487 AA.
ID ICP4_HSVB
AC P28925;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-FEB-1994 (Rel. 28, Last annotation update)
DE TRANS-ACTING TRANSCRIPTIONAL PROTEIN ICP4 (155 KDA IMMEDIATE-EARLY
DE PROTEIN).
DE IE OR 64.
OS Equine herpesvirus type 1 (strain Abdp) (EHV-1).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Alphaherpesvirinae; Varicelloviruses.
OX NCBI_TaxID=31520;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92295566; PubMed=1318606;
RA Telford E.A.R., Watson M.S., McBride K., Davison A.J.;
RT "The DNA sequence of equine herpesvirus-1.";
RL Virology 189:304-316(1992).
CC -1- FUNCTION: THIS IE PROTEIN IS A MULTIFUNCTIONAL PROTEIN CAPABLE
CC OF MIGRATING TO THE NUCLEUS, BINDING TO DNA, TRANS-ACTIVATING
CC OTHER VIRAL GENES, AND AUTOREGULATING ITS OWN SYNTHESIS.
CC -1- SUBCELLULAR LOCATION: NUCLEUS OF INFECTED CELLS.
CC -1- PTM: A LONG STRETCH OF SERINE RESIDUES MAY BE A MAJOR SITE OF
CC PHOSPHORYLATION.
CC -1- SIMILARITY: BELONGS TO THE ICP4/IE140/IE180 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M86664; AAB02515.1;
DR EMBL; M86664; AAB02499.1;
DR PIR; A36802; EDBE6.
DR HSSP; P04002; 1MFA.
DR TRANSFAC; T00401;
DR
DR Early protein; Transcription regulation; Trans-acting factor;
KW DNA-binding; Phosphorylation; Nuclear protein.
FT DOMAIN 181 213 SER-RICH.
FT FT 922 931 ARG/LYS-RICH (BASIC).
SQ SEQUENCE 1487 AA; 154868 MW; 1D4397838BD03680D CRC64;

```

Query Match 3.2%; Score 202; DB 1; Length 1487;  
 Best Local Similarity 19.9%; Pred. No. 0.019;  
 Matches 234; Conservative 134; Mismatches 455; Indels 350; Gaps 50;

```

QY 137 LRDQYSTERLSLSQSLQNLNKEGSEFYQYPAS---YHSNQTALGETTSPQLP--- 190
DB 16 IISNDF--GEDPLIRASA-----AEEGFTQPAAPDILYGSQNNFQVDDALSTPAVVI 67
QY 191 -----RGTQARATGQSFQGTTSRAGHLAGEPPAPPPPPPPPPPPPPPPPPPPPP 241
DB 68 PPPSPPEPRGGKAKRSPAAAGS-----GPPTPAAAQAQPSPA-----P 107
QY 242 DAPPAAMAAALYSSSTLPAPPRGSPPLAPOG-----SPTKLQRGSAFEGATY 292
DB 108 SPAPGIAALMKWHSVAVANGNGRRATGSSSPGGAADPVALDSTENCPSGPQEPFSS 167
QY 293 AAPRGSSPKQSPRLAKSTSTSPINI-----VSSAGLSPIRVTPPYQSTISSPHIQ 348
DB 168 ASGGGSPAPRVRSISISSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSDS 216
QY 349 LSTTIGTVAITLSPTRKLIVASSEQYKHSQELATY-ATLQRPGLAAGSAVASYSSQ----- 402
DB 217 DSDGEGEETPRP-----RHSGNAKRTSAAAGSPSSGGRPAAGATTPRSC 264

```

```

QY 403 -----HGLGPELRALQSPHHIDPIYEDRYQ----- 430
DB 265 RSGAASPCAPAPADASAPSPGCGILLPCARILE-----YLECVREANLAKTLERPEP 320
QY 431 -----KPPMRSLSSQSGDPLP-PAHTGYRTSTAPSSGCVDSVPLQRT--GSQHPQ----- 479
DB 321 AGMASPPRSPHRLPKQDPRKSAALAGASKRRRANRP-----IPOTQOAPAEAPQTAW 376
QY 480 -----NAAATFORASYAAGPASNADP---YRQLOCPSPVESYSGPALPEPTGLAR 531
DB 377 DLDMMNSQATGAAAAAASAPAAASCAPGVYQRPFLTPSGD--PWGSDP--PPMGVRY 433
QY 532 SPSSIDSIQKDRREFGWRDPELVYIOMLOHPSPVQSNAAAYLQHLGFGDNKTKAELRQ 591
DB 434 GGTGDS-----RDGLMDDEP---YVLAASRYAEQAQVPVFEVEMGDSYKQYNALV--- 481
QY 592 GGIQLVDLDLHRTVEYHRSACGALRNLYVGKANDDKIALKCGGIPALVRLKRTTDL 651
DB 482 -----RMVFESREMSWLN---SKLSGODONLQFCQKRTIAPR----- 518
QY 652 EIRELVTVGLWNLSGCDALKMPTIQQDALVLTNAVIIIPHSQWENSPLODDRKTIQLHSSQV 711
DB 519 GHGSEFTIGSVAN-----PLPHIGDMAAGNALMALPHAA--ASVAMSRIRYDQKSF 569
QY 712 LRNATGCLRNVSSAGEARRRRECDGLTDLALVYIQSALGSS-----EIDSKTV 761
DB 570 LQSLRRAYADMAVPRDEAGR---DSLAAVAGYPAQAAAAAASQOQPEAPASVRYREA 625
QY 762 ENCYCIL---RNLSTRLAETSGQGHMGDELDGLGCEANGKDAESSC---WGKKK 814
DB 626 YTRCAALGPRRKAAAAAAGTRAPRPSAFRLREL-----GDACVLACQAVFEALLR 678
QY 815 KKSQSDQMDGVP--LRPCAPPKGIOMLHPSTVYKRYLTLLSECSNPDLLEGAGALON 872
DB 679 LRGAASAVPGIDPELIPSPACPEAL-----CSNPAGLETAUSLYE 720
QY 873 LAAGSWKWSYTRAAVRKENGPLTVELLRIDNDRVCAVATLRNNALDVRNKKELIGY 932
DB 721 LRD-----LVERARLIGD----- 733
QY 933 AMRDLVHRLPGGNSNNTASKAMSDDTVTAVCTLHEVITIKRMBENAKALRDAGIEKLV 991
DB 734 --SDPTHTLL--GSDRLRAVRA---VLVYARTVAVLYVYNEGARARASAMTVTQAVF 784
QY 992 GISKSGDKHSPKYVKAASQVLNMQYRDLRLSKKDKGSOYHFVASSSTIERDQRPY 1051
DB 785 SIPLVGMGLGEAV-----SLAPPTRSQO---PSSSS---PGGEPR 820
QY 1052 SSSRT-----PSISPVRYSPNNRSASAPSPREMISLKER-----KTDYECTGSNAT 1098
DB 821 SGAASAEBSIQTLPLPMP--TVPGKQSAIVPSSHSQSPQHSQSGGAGATTATTCARATQ 878
QY 1099 YHAKGHEITSRKDMTQNTGISTLYLRNYSIGAPAEADIKHNOVSQAQVPQSPKDIETTY 1158
DB 879 NARSRGQOHQOKARKSPQ-AAASPAHLISOAMPSSSDDDRAIHGRPKGSKGRRS--EPLE 936
QY 1159 P-----FQNSTRYNDESFEDQVHHNRP 1182
DB 937 PAAQAGASASFSSSARKYDPSGPVDS-----PPA 965

```

RESULT 8  
 VRP1\_YEAST STANDARD; PRT; 817 AA.  
 ID VRP1\_YEAST  
 AC P37370; 006133;  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE VERPROLIN.  
 OS Saccharomyces cerevisiae (Baker's yeast).  
 GN VRP1 OR MDP2 OR END5 OR YLR337W OR I8300.13.  
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.  
 OX NCBI\_TaxID=4932;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RX STRAIN-A364;  
 RA MEDLINE=95058201; PubMed=7968536;  
 RT Donnelly S.F.H., Pocklington M.J., Pallota D., Orr E.;  
 RT "A proline-rich protein, verprolin, involved in cytoskeletal  
 RT organization and cellular growth in the yeast *Saccharomyces*  
 RT *cerevisiae*.";  
 RL Mol. Microbiol. 10:585-596(1993).  
 RN [2]  
 RC SEQUENCE FROM N.A.  
 RX STRAIN-S288C / AB972;  
 RA Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Du Z.,  
 RA Faveille A., Fulton L., Gattung S., Greco T., Kirsten J.,  
 RA Kuwaba T., Hallsworth K., Hawkins J., Hillier L., Jier M.,  
 RA Johnson D., Johnston L., Langston Y., Latreille P., Le T.,  
 RA Mardis E., Menezes S., Miller N., Nhan M., Pauley A., Peluso D.,  
 RA Rifken L., Riles L., Raich A., Trevaskis E., Vignati D.,  
 RA Wilcox L., Wohlman P., Vaudin M., Wilson R., Waterston R.;  
 RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: INVOLVED IN CYTOSKELETAL ORGANIZATION AND CELLULAR  
 CC GROWTH. MAY EXERT ITS EFFECTS ON THE CYTOSKELETON DIRECTLY, OR  
 CC INDIRECTLY VIA PROLINE-BINDING PROTEINS (E.G. PROFILIN) OR  
 CC PROTEINS POSSESSING SH3 DOMAINS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: Z26645; CAA81388.1; -;  
 DR EMBL: U19028; AAB67263.1; -;  
 DR PIR: S39626; S39626.  
 DR SGD: S0004329; VRPL.  
 KW Cytoskeleton; Repeat.  
 FT DOMAIN 5 14  
 FT 239 245 POLY-PRO.  
 FT DOMAIN 349 357 POLY-PRO.  
 FT DOMAIN 396 406 POLY-PRO.  
 FT DOMAIN 424 431 POLY-PRO.  
 FT DOMAIN 462 468 POLY-SER.  
 FT DOMAIN 704 708 POLY-PRO.  
 FT CONFLICT 308 308 P -> R (IN REF. 1).  
 FT CONFLICT 350 350 A -> R (IN REF. 1).  
 FT CONFLICT 689 689 V -> E (IN REF. 1).  
 FT CONFLICT 710 817 PSTMDGTGNSPSKRLKORLEFSTGSGTLQKHNTHTNOPDV  
 FT DVGRTTGGNSITGAGSKRRIYIDDSRFRKMNVSQMPD  
 FT RPRKRIKLYSPGSSVPLDITLFT -> HIRITPPLIA  
 FT FVKTLNNGYFLQVDRRCNTSITRIQIMMLA (IN REF.  
 FT 1).  
 SQ SEQUENCE 817 AA; 82593 MW; 24C752D5B1CA1C8 CRC64;  
 Query Match 3.28; Score 201; DB 1; Length 817;  
 Best Local Similarity 20.28; Pred. No. 0.0099;  
 Matches 154; Conservative 85; Mismatches 273; Indels 250; Gaps 34;  
 OY 6 PGAPPLGAMPYDPDSSASEK---TSSLSPGLNTSGDSEFTTSAIIASVKEQLQ 61  
 DB 150 PSSAP---PLPDPSSNAAPPIVPSPPAPPLPLSASAPKPYQNPMPVRAPHRS 205  
 OY 62 PERLTRELEAEROIIVASOLERCKLGSETGSSMSMAEEQOWOSQDGKDIEDLTGL 121  
 DB 206 HQKSSNLSLP-SVSAAPLPASLPTHSN----- 234  
 OY 122 ELVDSCLNSLQESGLDPODYSTGERPSLSLQSLQSKPEGSGFOYPASHSQTLALG 181  
 DB 235 -----PQAPPPPTPTIGLDSK---NIKPTDNAVSPSS----- 265

OY 182 ETTSPQLPARG-----TQARATGGSFSGTTS-----RAGHLAGEP-----APP----- 221  
 DB 266 ----SEVPAGGLEPFLAEINARSRSEGAIVGSTKIQTENHSPSQPPLPSAPPIPSH 321  
 OY 222 -PPPPPEPFAASLGSFAFLP---DAP-----PAAAMALYYSSSTLPAPPGSGPL 269  
 DB 322 APPLPPTAPPPPSLPVTSAPKKAISAPAPPPPLPAAAMSSASTNSVATPPPLAPPL 381  
 OY 270 AAPQGSPTKLRGSGAPGATYAAPRGSSPKQSSRLAKSTSTSPINIVYSSAGLSPI 329  
 DB 382 PNTTSVPNPK---ASSMP-----APPPPPPPPGASTSSALS---ASSIPLAPL 425  
 OY 330 RVTSPTVOSTTSSPIHQLSSTIGYATLSPTRKRLVHNSHQSHSOELVATATL----- 385  
 DB 426 PPPPPSVATVPSAP-----PPPTLTITNKPS-----ASSQSKTSSSSSSANTPPGP 475  
 OY 386 -----QRPGLAAGSRASYSGHG-----LQBELRALQSP 416  
 DB 476 LPFLAEIQKKRDRFVVGDTGYTDQKQEDYIGSSKQDNVBPSPISINPPKQSSONG 535  
 OY 417 EHHIDPIEDRYQK-----PPKRSLSQSGDPLPA-----HT 450  
 DB 536 MSFLDEI-BSKLRKQTSNANFAPPH--TDMAPPLPSAPPPPTSLPTASGDHT 592  
 OY 451 GT-----YRTSAPSSPG-VDSVPLQRTGSGHQGNAAATFORASYAAGPASN----- 499  
 DB 593 NDKSEVYLGKKKAKAPALGCHVPPPVYLDSDSKNNVPAASL---LHDVLPSSNLEKP 649  
 OY 500 -ADPYRQLOVCPSESP-----YKSGPALPPEG-TLARSIDSIOKDPREFGRQDE 551  
 DB 650 PSPPVAAAPPLPFPASAPLPQOOSVSTSISSPPPVATLSVRETETESISKNPTR-SPPPP 708  
 OY 552 LPEVIGMLQHQPSPVQSSNAAYLQHLCEFGDNKIKAEIRROGGIQLVLDLDRM----- 605  
 DB 709 SPSTMDTGTSSNPS-----KNIKQRLFTSG-----STLQKHNTHTNQ 747  
 OY 606 -TEVHRSACALNRLVYKAN-----DNNKIALKNCGIP 639  
 DB 748 PDVDVGRYTTIGSGNSIVGAKSGNERIYIDDSRKRMNVNSQMP 789  
 RESULT 9  
 SMPL\_HUMAN STANDARD; PRT; 1902 AA.  
 ID 014487; Q9UPZ1.  
 DT 01-OCT-2000 (rel. 40. Created)  
 DT 01-OCT-2000 (rel. 40. Last sequence update)  
 DT 01-OCT-2000 (rel. 40. Last annotation update)  
 DE SWI/SNF-RELATED, MATRIX-ASSOCIATED, ACTIN-DEPENDENT REGULATOR OF  
 DE CHROMATIN SUBFAMILY F MEMBER 1 (SWI-SNF COMPLEX PROTEIN P270) (B120).  
 GN SMARCE1 OR CIOF4.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., AND MUTAGENESIS.  
 RX MEDLINE=20221560; PubMed=10757798;  
 RA Dallas P.B., Pachione S., Wilsker D., Bowrin V., Kobayashi R.,  
 RA Moran E.;  
 RT "The human SWI-SNF complex protein p270 is an ARID family member with  
 RT non-human specific DNA binding activity.";  
 RL Mol. Cell. Biol. 20:3137-3146(2000).  
 RN [2]  
 RP SEQUENCE OF 1-1175 FROM N.A.  
 RX MEDLINE=98094256; PubMed=9434167;  
 RA Takeuchi T., Chen B.-K., Qiu Y., Sonobe H., Ohtsuki Y.;  
 RT "Molecular cloning and expression of a novel human cDNA containing CAG  
 RT repeats.";  
 RL Gene 204:71-77(1997).  
 RN [3]  
 RP SEQUENCE OF 1-1132 FROM N.A.





OY 1113 MTAONTGISLYRNSYCAPADIKHNOVSAQPYOEPSRKDYETOPFONSTRNDE 1169  
 DB 981 VESYSEDESKF-CSYGOYPADLAHKITHSANHMD-NDGELDPPIINSLKYSDE 1032

## RESULT 11

IE18\_PRIVIF STANDARD; PRT; 1461 AA.

ID IE18\_PRIVIF

AC P11675;

DT 01-OCT-1989 (Rel. 12, Created)

DT 01-APR-1990 (Rel. 14, Last sequence update)

DT 01-FEB-1994 (Rel. 28, Last annotation update)

DE IMMEDIATE-EARLY PROTEIN IE180.

GN IE.

OS Pseudorabies virus (strain Indiana-Funkhauser / Becker) (PRV).

OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;

OC Alphaherpesvirinae; Varicelloviruses.

OX NCBI\_TaxID=31523;

RN [1]

SEQUENCE FROM N.A.

RA MEDLINE=89315207; PubMed=2546124;

RT "DNA nucleotide sequence analysis of the immediate-early gene of

RT pseudorabies virus".

RT Nucleic Acids Res. 17:4637-4646(1989).

RN [2]

REVISED.

RA Cheung A.K.;

RL Submitted (NOV-1989) to the EMBL/Genbank/DBJ databases.

CC -1- FUNCTION: THIS IE PROTEIN IS A MULTIFUNCTIONAL PROTEIN CAPABLE

CC OF MIGRATING TO THE NUCLEUS, BINDING TO DNA, TRANS-ACTIVATING

CC OTHER VIRAL GENES, AND AUTOREGULATING ITS OWN SYNTHESIS.

CC -1- SUBCELLULAR LOCATION: NUCLEUS OF INFECTED CELLS.

CC -1- PTM: A LONG STRETCH OF SERINE RESIDUES MAY BE A MAJOR SITE OF

CC PHOSPHORYLATION.

CC -1- SIMILARITY: BELONGS TO THE ICP4/IE140/IE180 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>

CC or send an email to [license@sdb.ch](mailto:license@sdb.ch)).

CC -----

DR EMBL; X15120; CAA33214.1; .

DR PIR; S04713; EDBE1F.

KW Early protein; Transcription regulation; Trans-acting factor;

KW DNA-binding; Phosphorylation; Nuclear protein.

FT DOMAIN 390 405 POLY-SER.

FT SEQUENCE 1461 AA; 149833 MW; 7F31E7ABE403B208 CRC64;

Query Match 3.1%; Score 197; DB 1; Length 1461;

Best Local Similarity 20.9%; Pred. No. 0.03;

Matches 177; Conservative 99; Mismatches 347; Indels 224; Gaps 37;

OY 1 MFARKPPGAPLIGAMPVPDPSPASSEKTS-----LSPGLTNSGDSGETTETATLAS 54

DB 18 LIAAAAAAEEEGIASGPGGSGSRRSSGDELLFGRG-GLFSDDAAEAL-AAVILAA 74

OY 55 VAEQELQFRLRLREAEQIYASOLERCKLGESESSSSSSAAEPOFOWOSQDQKIE 114

DB 75 A-----AGATRRPPPPSAOORHARRGS--GEIYVLDDEDEE-----E 110

OY 115 DELTGLLELVDCISLSQSSG-----ILDPOYS--TGERPSILSQAOLNSKPEG-- 164

DB 111 DE--PESPAAGSPGRALHSGSHGHLVIGPFRASGPPPPALAAAEAGAPGPGRS 168

OY 165 --SFQYPASYHSNQTALGETTTPSQLPARGTQARATGOSFSOGTTSSRAGHLAEPAPPP 222

DB 169 SPASAPASSSSSGSSGSSGSPGSAAPRRMSPAR-----GDPVGERP 211

OY 223 PPPPREPAPSLGSAFLHPDAPAAAAALYSSSTLAPPGSGSPLAPOGGSPTKLOR 282

DB 212 AARPPRP-----APPAQPAVAAPARRGAPASAPAG-PVASAP-----G 251

OY 283 GGAPEGATYAAPRGSSPKQSPSLAKSYSTSPINIVVSSAGLSPIVTSPTVOSTIS 342

DB 252 GGGAPSG--GGDRGRHHQHREPL-----LDEPAAARRLDP--RPLARSEVS 295

OY 343 SPTIQLSSTIGTYTTLSPYKRLVHASQYKSHQSELYATLOR-----PGSLAAG 394

DB 296 SNPNSSSSST--TVAAYEPVARGPEKDEDGLAGD--GAPLQROPRRRARGCALRG 351

OY 395 SRASYSQHGLHPELRLALQSEHHIIDIYEDRYQKPMKSLQSGQDPLPAHTGYR 454

DB 352 -RGFSSSSGSGSDSLSPARSPAPRAAAAAARRASSSSSSSSS----- 398

OY 455 TSTAPSPGVD-----SVPLQRTGSQHGP-----ONAAATFORASYAAGPASN 498

DB 399 SSSSSSSGGEDEGVAPGAPLAPAGPPSPAPAPAPASASATSSSAAAPAPAPE 458

OY 499 YADPTROLQYCPG-----VESPTKSGPALPPEGT--LARSPTDSTQK--DPREF 545

DB 459 PARPPRRRRRSTNNHLSLMADGPPPTDGLPLTPIGEPWPGSDPADGRVRYGAGDSREG 518

OY 546 GWRDELEPEVIOMLQHOPPSVSNAAAYLOHLCFGDNKIKAEIRGGIQLLVLLDHRM 605

DB 519 LMDELD-----VRQAAARKRAAAGPVVFTIPRM--GDSKQHE----- 554

OY 606 TEVRSACALRNIVYGRANDNKAINKCGIPALVRLKTTDLIRE----- 655

DB 555 -----ALVRLYSGAGEAMSWLON--PRM-----QANDQFNFQCRVHAPH 596

OY 656 ----LYTGVLNLSGCDALKMPITQDALAVITNVTIIPHSQWENSPLODDRKIOLHSSQV 711

DB 597 GHGSETTSV-----TPPLPHIGDAAQDPLMAAPHA--VSAAVMSRRYRTQKTFI 647

OY 712 LRNATGICLRNYSAGEARRRRRECDDGLTDLALVYIOSALGSSP--IDSKTVENCICIL 768

DB 648 LOSLRRAADADAYGRRADPAPG--ATVEALCARVRAAPAAQPGVRELADACVLAC 705

OY 769 RNLSTRL 775

DB 706 RGVLERL 712

## RESULT 12

YM96\_YEAST

ID YM96\_YEAST

AC 004893;

DT 01-NOV-1997 (Rel. 35, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 01-NOV-1997 (Rel. 35, Last annotation update)

DE HYPOTHEMETICAL 113.1 KDA PROTEIN IN PRES-FET4 INTERGENIC REGION.

GN YMK317W OR YM9924.09.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.

OX NCBI\_TaxID=4932;

RN [1]

SEQUENCE FROM N.A.

RA STRAIN=S288C / AB972;

RA Churcher C.M., Louis E.J., Barrell B.G., Raeburn M.A., Walsh S.V.;

RL Submitted (NOV-1995) to the EMBL/Genbank/DBJ databases.

CC -1- DOMAIN: CONTAINS MANY SER/THR-RICH DOMAIN AND REPEATS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>

or send an email to license@isb-sib.ch).

CC -----  
 CC EMBL; Z54141; CA90835.1; -  
 DR SCD; S0004936; YMR317W.  
 KW Hypothetical protein; Repeat.  
 SQ SEQUENCE 1140 AA; 113070 MW; 0153EBCA24FE5427 CRC64;

Query Match 3.1%; Score 195; DB 1; Length 1140;  
 Best Local Similarity 19.4%; Pred. No. 0.029;  
 Matches 242; Conservative 177; Mismatches 505; Indels 322; Gaps 52;

20 QPSASAKTSLSL-----SPGLNTSNGDGET-----ETTSAILASYKEDQLG 62  
 128 QLSSTETSTISSAIAQTSSPOTSSNGGSSSEPLGKSSVLETTAKSSDTPAVTSST 187  
 63 ERLRELEAEQOIVASOLEKCLGSEFSMSMSMAEFOQOMODQKIDELTTLGLE 122  
 188 TLTLL-DYSSSPKISS-----GSATSVGTISDASK-----E 218  
 123 LVDSCLRSLOESGILDPODYSTGERPRLSQALQNSKPEGSFOYPASTHSNOTLALGE 182  
 219 VFSS-----STSDVSSLS-----STSSPASSTISETLPFSS 250  
 183 TTPS--QLPARGTQARATGGSFGSGTTSRAGHLAGEPAPPPPPPPPPPPPPPPPPPP 240  
 251 TLTSTSSVSSSEAPSAITSSVSSSEASSSTSSVSE--APLATISSVSSSEAPSTSSVS 309  
 241 PDAPPAALAAALYSSSTLPAPPPGSPPLAPOGSPFKLRGGSAPGATYAAPRGSSP 300  
 310 SEASSSTSSVSSSEISSTSSVSSSEAPLA-----TSSVSEAPSTSSVSSSEISS 362  
 301 KQSPRLAKS-YSTSSPLNIYSSAGLSPIRTVSPPTVOSTISSPFIHOLSTRTGTATL 359  
 363 TTSSVSEAPLATISSVSSSEAPSTSSVSEAPSTSSVSEAPSTSSVS--SET 420  
 360 SPTRLVASQYKSHSOELATATL--QRPGSLAA-GSRASYSQHGHLGPELRALQS 415  
 421 SSTSSVSSSEVSS-----ATSSIVSEAPSAISSSLASRLFSKNTVSTSLVATEA 473  
 416 PEHHIDPIEDRYOKPRMSLSOSQGDPLPAHTGTYRTSTAPSSGCVDSVPLQRTGSO 475  
 474 -----SSVTSLSLRPSSETLNLSNITSSLS 498  
 476 HGPNMAAATFORASYAGPASNYADPYRQLYCPSEVSPKSGPALPPGTLARPSI 535  
 499 TGVNYSVSTTSAASSTIGSVSSSNMAMATSKTSTSSLSKSVIFGNSVYTTSPSA 558  
 536 D-SIOKDPREFGWRDPELVYIQLHOFPSVQSNAAAYIQLHCFGNKIKAEIRROGI 594  
 559 SISLAPSLPVSMD-----ITSSSEASSISNLS--SSAPSDNNSTI-----ASA 602  
 555 QLLVDLDRHRTVEVHRSACGL-RNLYVGKANDDNKIALKNCGI--PALVRLLRKTTDL 651  
 603 SLIV-----TKTKNSVSVSIVSITSESTNENSLATSTSLSKMKAFAKSLSTNAT 655  
 652 EIRELVTGLVNLNLSGCDLKKPIIODALAVLT-NAVITPHS--GM-----ENSLQDRK 703  
 656 SASNPPTGFSMSHRTVITPGSTSSASLAISVYSSSLAISTPSPSSPTT----- 711  
 704 TQLHSSQVLRNATGCLRNVSSA-----GEEARRRMRCDGLTALLVYIQSALGSSEID 757  
 712 -----STLVTSAPSTVSSMTSAPFINNSTSAPSTASTESTSSISSVPLASGVT 767  
 758 SKTVANCICILRNLSYRLAETSGQHGTELDGLGCEANGDAESGQWKKKKKK 817  
 768 SSLAAH-----NLTFSPSTSSAQIV-----SKSTSSSTI----- 798  
 818 SODQMDGVGLPDCAEPPKGIQMLHPSIVKPYLTLSECSNPPTLLEGAAGALON----- 872  
 799 -----LVMP--RIDRSGNNSPASTSIATSLNKTTFF 827  
 873 --LAAGSKMSVYIRAAVRKEKGLIVLELLRID--NDRVVCAYATALRNALDV---RN 925

DB 828 SLSSTS-----AHRNIFNSVTLATAKQIEFLINSTVCSNPTPNYNTKTVISRE 879  
 QY 926 KELICKYMRDLVHRLPGC--NNSNNTASKMSP-DIVTAVCCILHEVITKMNMAALR 982  
 DB 880 TTAIG-----TVNISCSCGCTKRNKSTLLITIDASTVTTCEKEV----- 921  
 QY 983 DAGGTEKLVGISKSGD--KHSPKVKASQVLSNMQYRDLRSILY--KKDWSQYHFA 1038  
 DB 922 -----TSTISGDAEHTTSTKISNFTSPSEFKMKISQETKRAKPPSEIVRS 971  
 QY 1039 SSSITDRORPYSSTRTSPISPVVSPNNRSASAPSPREMIKREKTDYECTGSNAT 1098  
 DB 972 SSSFEVK-----TSPTTKASP-STSPSEKKAAGNTSVATNAPSPSTSPSQGTGSTSV 1023  
 QY 1099 YGAKGEHTSKRDAMT---AQTGISTITRNSYCAPADIDKHQVSAPOVQEPSRKDY 1155  
 DB 1024 -EGAKSKSTKNSGVSSTKAKNT--STVAKSSTESP--IGRGTELTIVS--SQKSL 1076  
 QY 1156 TYQPPQNSTRNDESEFED--QVHHRPPASE---YTMHLGLKSTG 1195  
 DB 1077 TSC-LSSTEKYKNTTTPTAIHGTSSAKOSTIYIVSTAKENTG 1121

RESULT 13  
 ID APC\_RAT STANDARD: PRT: 2842 AA.  
 AC P70478;  
 DT 01-OCT-2000 (Rel. 40, Created)  
 DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE ADENOMATOUS POLYPOSIS COLI PROTEIN (APC PROTEIN).  
 GN APC.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FSCHER 344/N; TISSUE=Brain;  
 RX MEDLINE=96116966; PubMed=8563176;  
 RA Toyota M., Ushijima T., Kakiuchi H., Watanabe M., Imai K., Yachi A., Sugimura T., Nagao M.;  
 RA cDNA Cloning of the rat APC gene and assignment to chromosome 18.;  
 RL Mamm. Genome 6:746-748(1995).  
 RP [2]  
 RP MUTAGENESIS.  
 RC STRAIN=SPRAGUE-DAWLEY, AND FISCHER 344/N;  
 RX MEDLINE=95148647; PubMed=7846077;  
 RA Kakiuchi H., Watanabe M., Ushijima T., Toyota M., Imai K., Weisburger J.H., Sugimura T., Nagao M.;  
 RT Specific 5'-GGGA-3'->5'-GGA-3' mutation of the APC gene in rat colon tumors induced by 2-amino-1-methyl-6-phenylimidazo[4,5-b]pyridine.;  
 RL Proc. Natl. Acad. Sci. U.S.A. 92:910-914(1995).  
 CC -1- FUNCTION: TUMOR SUPPRESSOR. ALLOWS THE RAPID TURNOVER OF BETA-CATENIN. APC ACTIVITY CORRELATED WITH ITS PHOSPHORYLATION STATE ALLOWS THE DOWNREGULATION OF CYTOPLASMIC BETA-CATENIN (BY SIMILARITY).  
 CC -1- SUBUNIT: FORMS HOMODIMERS AND ASSOCIATES WITH CATENINS (BY SIMILARITY).  
 CC -1- PTM: PHOSPHORYLATED BY GSK-3B (BY SIMILARITY).  
 CC -1- SIMILARITY: CONTAINS 7 ARM REPEATS.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL; D38629; BAA07609.1; -  
 CR HSSP; Q02248; 2BCI.



InterPro: IPR000225; -  
 DR Pfam: PF00514; Armadillo\_seg; 4.  
 DR PROSITE: PS0176; ARM\_REPEAT; 1.  
 KW Anti-oncogene; Phosphorylation; Coiled coil; Repeat.  
 FT DOMAIN 1 728  
 FT DOMAIN 1 62  
 FT DOMAIN 125 260  
 FT REPEAT 451 493  
 FT REPEAT 503 545  
 FT REPEAT 546 589  
 FT REPEAT 590 636  
 FT REPEAT 637 681  
 FT REPEAT 682 723  
 FT REPEAT 724 765  
 FT DOMAIN 739 2831  
 FT DOMAIN 1130 1155  
 FT DOMAIN 1556 1575  
 FT DOMAIN 1864 1891  
 FT MUTAGEN 523 523  
 SQ SEQUENCE 2842 AA: 310530 MW: 308284834E8F47 CRC64;  
 C->R: IN AN IO-INDUCED COLON TUMOR.  
 HIGHLY CHARGED.

Query Match 3.1%; Score 194.5; DB 1; Length 2842;  
 Best Local Similarity 18.1%; Pred. No. 0.096;  
 Matches 230; Conservative 166; Mismatches 424; Indels 451; Gaps 50;

23 SAEKTSLSLP-----GLTSSGDSSETE-TTSALISVKEOE-----IQ 61  
 100 SVSRSGECSPPVPMGSPRRAPVNGSRETYLELEKERSLLADLDEEKKDWYQAQ 159  
 62 FERLFELEA-----EROIVASOLE-----RCKLGSETGSSMSAEEQOMQS 106  
 160 LQMLTRIDLPITENFSIQDTMRQLEEARQITAAEBQIGTQDMKRAQRIARI 219  
 107 QDGKQIDELTGLLELVDSICRSIQESGILDPQDYSTERPSLSQSALQINSKEGSE 166  
 220 QQTEKDI-----LRVROLLOS-----QAEEAERSS-----QSKHE-- 249  
 167 QVPASVHNSQTLALETTSQLPARKTOARATQSGTSQGTSGAGHLAPPEAPPPPPP 226  
 250 --TASHEAEKOLEGGVAAASNLTATSGS-----GQSSAAKRVH----- 284  
 227 REPFAISLGAFLPAPAPAAAALYSSSTLPAPPPGSPPLAQQSGPTKLRQGSAA 286  
 285 -----ETAGVLSSSGHSHAPRLTSHLG-----TKVENVYS- 315  
 287 PEGATTAAPRGSSPKOSPSRLAKSVSTSSPINIVVSSAGLSPIRYVTSPTVOSTISSPI 346  
 316 -----LLSMIGTHDKDMRSTLLAMSSQDSCISMROSCCLPLLI----- 355  
 347 HOLSTIGVYATISPTKRLVHASEQYSKSHSQELIYATATLQIRGSLAAGSRASYSSQHG-- 404  
 356 -----QLHNGDKDS-----VLGNSRSGSEKARARAS-AALHNII 389  
 405 HLGPE-----LRAQSPENHIDPIYEDRVYOKPMSRSLSSQSGPPLPAHGTGR 454  
 390 HSGPDCKRGRREIRVHLLEQIRAYCETCWE---WQEAHEGMDCK-NPM- 437  
 455 TSTAESSPEVDVPLQRTGSSQHGPNAAAATFORASAYAGPANYADPYROLQYCPSVES 514  
 438 -----APVEH-----QICPAV-- 448  
 515 PYSKSGPALPREGTLARSPSIDSIQKP-RERGWMDPELPEYIQMQ-----HQ 562  
 449 -----CYLMKLSFDEERHANNELG---GLQALAEILOYDCEMHGLTDHY 491  
 563 PFSVSNAAAYLOHLCEFGDNKIKAEI-RRQGIQLLVLDLDRHMEVRSACALRNLYV 621  
 492 SYTLRRYAGMALNLTFQGVANKATLCSMKGCMRALVADQKSESDIQOVIAYVLRNLSW 551  
 622 GRANDNKIALKNCGGIPLVLRLLKTTDLTRELVTGVANLSS-CDALKMPT--TODA 678  
 552 -RADVNSKKTTLREVGSVKALMECALELVKKESTKLSVLSALNWLNSHCHENKADICAVGDA 610

679 LAVLTNAVITPHSGWENSLPQDRKIQLOHSSQVLRNATGCLRNVSS---AGEPARRRRE 735  
 611 LAFIVGTILY-----RSQNTLAIISGGGILRNSSSLIATEDHRQILRE 656  
 736 CDGLTDAIIVYQALGSSEIDSKTVENCYCLRNLSYRLAETSSQGMGTDELGLIC 795  
 657 NNCLQTLQHLKSHSL-----TIVSNACGTLNLSAR----- 688  
 796 GEANKRDAESSGCMKKKKKKKSDQMDGVPLPDCAPPPGIOMLWHPSTVYKPYLTLS 855  
 689 ---NPKDEA-----LMDGAVS-MLNKLI 709  
 856 ECSNPDTLEGAALONLAA-----GSMKSYIIRAVKEREGLPIVELL 901  
 710 HSKHMMIMGSAALRNLMANRPAYKKNANIMSPGSSLSPLH---YRKQALAEIDAQ 765  
 902 -----RIDNDRVCAVATLRNMLDVNRKELIGKYAMBOLVHRLPGNNSNNTASKA 954  
 766 HLETFEDNIDN-----LSPKASHRSKORHKONLYGDVYFPASRH---DDRRSDN- 811  
 955 MSDDTVTAVCCITHEVITKNNMENAKALRDAGGIEKLVGISKSG----- 998  
 812 PFTGMMTVLSPILNTTVLPSSSSSRGSLDSSRSEKDRSLEREKIGLSTYHSATENPGTS 871  
 999 DKHSPPVKAASQVLSNMQYRDLSLYKKDQWS-----QYHFVAASSSTTERDQRPYSS 1053  
 872 SKRGQLSATQAQIAKVMEEVSALHT--SQDDRSPASAAELHCVAE---ERTARRSSA 925  
 1054 SRTPTISPVYRSPNNRKSASAPREMI SLKERTIDVCTSNATY-----HGANGEH 1106  
 926 SHT-----HPNTHNPAKSESSNRCOSMPYAKVEYKRSSNDLSNVTSSDYGKKGOM 977  
 1107 TSKDAMTAQONTGISTLYRNSYGAPAEIDIKHNOYSAQVPPQEPKRYETVQPPQNSTRN 1166  
 978 KPVSYSYSEDEGKFC-----SYGQYPADLAKIHSAHMDNGELD-----TPIN 1024  
 1167 YDESFEEDQVH 1177  
 1025 YSLKYSDEQLN 1035

RESULT 14  
 YHLL\_EBV  
 ID YHLL\_EBV STANDARD; PRT; 660 AA.  
 AC P03181;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE HYPOTHETICAL BHLFI PROTEIN.  
 OS Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).  
 OC Viruses; dsDNA viruses; no RNA stage; Herpesviridae;  
 CC Gammaherpesvirinae; Lymphocryptovirus.  
 OX NCBI\_TaxID=10377;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=84270667; Pubmed=6087149;  
 RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,  
 RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seglin C.,  
 RA Tuffnell P.S., Barrell B.G.,  
 RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome."  
 RL Nature 310:207-211(1984).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 DR EMBL: V01555; -; NOT\_ANNOTATED\_CDS.  
 DR PIR: A03742; Q0BE3.

Hypothetical protein; Early protein; Repeat.  
 FT DOMAIN 149 648 4 x 125 AA TANDEM REPEATS.  
 FT REPEAT 149 273 1.  
 FT REPEAT 274 398 2.  
 FT REPEAT 399 523 3.  
 FT REPEAT 524 648 4.  
 SO SEQUENCE 660 AA; 66244 MW; 86DA1D67A37152A2 CRC64;

Query Match 3.0%; Score 194; DB 1; Length 660;  
 Best Local Similarity 22.4%; Pred. No. 0.016;  
 Matches 161; Conservative 48; Mismatches 271; Indels 240; Gaps 33;

7 PGAAPLCAMPYPPDPSSASEKTSLSPLGLTSMNGSEFETTAIIA-SYKEDLEQERL 65  
 11 PRTPLPHCPPLPGAPDOOTRLPPGW-----GQRTAPTOVGLADAASPELDQAS 64  
 66 TRELEAEROIYASOLECKIGSETGSSMSAEEQFOMOGOKDIEDELTTGLEIVD 125  
 65 GARGGGNRVAG--KGRGTAPRSQSRITGPAE-----QADHANSPTG----- 108  
 126 SCIRSLQESGILDPO-----DYSTGE-----RPSLSQ----- 153  
 109 -----GCSDPQRSPRTROAGYALGEGSAGIGSRPPHAFQVOMGARNPGCPTW 159  
 154 ---SALQLNKREPSFYQYPASYSNQTALGETTPSOLPARGTQARTGSGFSGTTSRA 210  
 160 RRRGAGRGHPPGAGRGPTGGRAPAPGCTPAAPG---GGAIVSGATPHR 214  
 211 GHLAGPERAPP-----PPRREFAP-SLGSAPHLDPAP-----AAAAALY----- 254  
 215 ER--GSGADPAPPAARLPREHGEPLRLQDLAAQRCPAGPPRTSGAAQRTHRPPGCP 272  
 255 -----SSSTLPAPRGSPPLAAPGGSGPTLQRGSA--P 267  
 273 RSARNPGCPRTWRRRSGAQRCHPPGAGQRPSPGTPGAPPAAPGAPGGAAPV 332  
 288 EGATYAAPRGSPKPSRSLAKSYSTSSPITIVVSAGLSPIRTYTSPTVQSTISSPIH 347  
 333 SGATPHPRGSGAPDP-----AAARLPERR-QEERLQ----- 365  
 348 QLSSTIGTYATLSPTRKLVHASEQYSKHSELYATLQRPGLAAGSRASYSQSHGLG 407  
 366 DLAAQRCPAGPPRTSGAAQR---THRRPGCPRSARNPGCPTWRRRS-GAQRGHP 421  
 408 P-----ELRALQSPRHHDITIEDRVYQKP 433  
 422 PGAGRPSPGTPGAPPAAPGAPGGAAPVSGATPHPRGSGAPDPAPPAARLP 481  
 434 MR-----SLSOSQGDPL--PPAHTG--TYRTSTAP-----SPGVDSPVLQRTGSQ 475  
 482 EROEPLRLQDLAAQRCPAGPPRTSGAAQRTHRRPGCPRSARNPGCPTWRRRSQAQ 541  
 476 --HGFQNA-----AAATFORASVYAGPASNADYRLOLQCPSESYSK--S 519  
 542 RGHPPPGAGORPSGTGGRPAAPGAPGTPPAAPGPGGAA-----VPSGATPHPRGS 593  
 520 GPALPPECTLARSPSIDIQDPRFGWDEPLEVIONLQH--QPPSYNSNAAYLQIH 576  
 554 GPADPP--AAARLP-----PER--QEERLQDLAAQRCPAGPPRTSGAAQRTH 640

## RESULT 15

SY11\_HUMAN STANDARD: PRT: 1575 AA.

AC 043426; 043425;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DE 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE SYNAPTOJANIN 1 (EC 3.1.3.56) (SYNAPTIC INOSITOL-1,4,5-TRISPHOSPHATE 5-  
 PHOSPHATASE 1).  
 GN SYMOL.  
 OS Homo sapiens (Human).

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 CC NCBI\_TaxId=9606;  
 CC (1)  
 CC SEQUENCE FROM N.A.  
 CC TISSUE=Cerebellum;  
 CC MEDLINE=96088905; PubMed=9428629;  
 CC Haffner C., Takel K., Chen H., Ringstad N., Hudson A., Butler M.H.,  
 CC Salcini A.E., Di Fiore P.P., De Camilli P.;  
 CC "Synaptojanin 1: localization on coated endocytic intermediates in  
 CC nerve terminals and interaction of its 170 kDa isoform with Eps15.";  
 CC FEBS Lett. 419:175-180(1997).  
 CC - FUNCTION: INOSITOL 5-PHOSPHATASE WHICH HAS A ROLE IN CLATHRIN-  
 CC MEDIATED ENDOCYTOSIS.  
 CC - CATALYTIC ACTIVITY: D-MYO-INOSITOL 1,4,5-TRISPHOSPHATE + H(2)O =  
 CC D-MYO-INOSITOL 1,4-BISPHOSPHATE + PHOSPHATE.  
 CC - ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG ISOFORM/SYNAPTOJANIN-  
 CC 170 (SHOWN HERE) AND A SHORT ISOFORM/SYNAPTOJANIN-145; ARE  
 CC PRODUCED BY ALTERNATIVE SPLICING.  
 CC - TISSUE SPECIFICITY: CONCENTRATED AT CLATHRIN-COATED ENDOCYTIC  
 CC INTERMEDIATES IN NERVE TERMINALS. THE LONG ISOFORM IS MORE  
 CC ENRICHED THAN THE SHORT ISOFORM IN DEVELOPING BRAIN AS WELL AS  
 CC NON-NEURONAL CELLS. THE SHORT ISOFORM IS VERY ABUNDANT IN NERVE  
 CC TERMINALS.  
 CC - DOMAIN: BINDS TO EPS15 (A CLATHRIN COAT-ASSOCIATED PROTEIN) VIA A  
 CC C-TERMINAL DOMAIN CONTAINING THREE ASN-PRO-PHE (NPF) REPEATS.  
 CC - DOMAIN: THE C-TERMINAL PROLINE-RICH REGION MEDIATES BINDING TO A  
 CC VARIETY OF SH3 DOMAIN-CONTAINING PROTEINS INCLUDING AMPHIPHYSIN,  
 CC SH3P4 AND GRB2.  
 CC - SIMILARITY: IN THE CENTRAL SECTION; BELONGS TO THE INOSITOL-1,4,5-  
 CC TRISPHOSPHATE 5-PHOSPHATASE FAMILY.  
 CC - SIMILARITY: CONTAINS 1 SAC1 DOMAIN.  
 CC - SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).  
 CC -----  
 CC This SWISS-Prot entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC EMBL: AF009040; AAC51922.1; -  
 CC EMBL: AF009039; AAC51921.1; -  
 CC MIM: 604297; -  
 CC InterPro: IPR000300; -  
 CC DR Pfam: PF00783; IPIC: 1.  
 CC DR PROSITE: PSS0102; RRM: 1.  
 CC KW Hydrolyase; Alternative splicing; Repeat; Endocytosis; RNA-binding;  
 CC Multigene family.  
 CC FT DOMAIN 1 499  
 CC FT DOMAIN 500 899  
 CC FT DOMAIN 900 1575  
 CC FT DOMAIN 902 971  
 CC FT DOMAIN 903 1036  
 CC FT DOMAIN 1033 1036  
 CC FT DOMAIN 1108 1113  
 CC FT DOMAIN 1113 1129  
 CC FT DOMAIN 1126 1129  
 CC FT DOMAIN 1487 1490  
 CC FT DOMAIN 1487 1490  
 CC FT DOMAIN 1540 1546  
 CC FT DOMAIN 1396 1419  
 CC FT DOMAIN 1396 1398  
 CC FT REPEAT 1396 1408  
 CC FT REPEAT 1408 1417  
 CC FT REPEAT 1417 1419  
 CC FT REPEAT 1419 1419  
 CC FT VARSPLIC 1306 1311  
 CC FT VARSPLIC 1312 1575  
 CC SO SEQUENCE 1575 AA; 173345 MW; 50646F6CC043B9E7 CRC64;

Query Match 3.0%; Score 194; DB 1; Length 1575;  
 Best Local Similarity 24.1%; Pred. No. 0.049;  
 Matches 137; Conservative 56; Mismatches 221; Indels 154; Gaps 29;

QY 105 QSDGQKDIDELTTGLELVDSIRSLQESGIL-----DPQDYSTERPSL 150

```
Db 969 KSPDWIKNLEEMMS--LEKISIALPSSSTSTLLGEDAEVAADFMEGDDVDDYS-AEVEEL 1025
QY 151 LSO-----SALQNSKPEGSFOYPASYNQTLAIGETTPSQLPARGTQA--RATGQSF 203
Db 1026 LPOHLQPSSSSSGLGTSPESS--PRT-SPCQSPITISEGVPVSLPIRPSRAKSPRTGPPSA 1081
QY 204 QGT---TSRAGHLAGPEPA-----PPPPPPPREPPAPSLGSAFHLDPADPAAAAAALYSS 256
Db 1082 QSSPIDAQAPATPLPQKDPQAQLEPKRPPPPR-----PVAPP----- 1117
QY 257 STLPAAPPGGSPPLAAPQGGSPTKLORGG-SAPEGATYAAPRGSSPKQSPSRLAK---SYS 312
Db 1118 -TRAPAPQRPAPPSPGASRPATRKKEFGIGAPPSPGVARRMEAPK-SPGTTKRDNIGRS 1175
QY 313 TSSPITNIV--SSAGLPIRVTSPTVQSTISSPIHQLSSTIGTYATLSPTKRLVHASE 370
Db 1176 QPSFOAGIAGPGPAGYSTARTTIPPRAGVYISAPQSHARASA-----GRLTP-----E 1222
QY 371 QYKHSQELVATLQRPGLAGSRASYSQHGHLGPELRAIOSPEHHIDPIYED---- 426
Db 1223 SQSTSETSKSGSTFLPEP---LKPQAAPPQ--SLPPPAQRLQEP--LVPAAPMPQS 1274
QY 427 -----RYOKPPMRSLQSOGDPLPPAHTGTYRTSTAPSPGVDSVPLQRTGSQHGPN 480
Db 1275 GPQNLFTPPQPPPRSRSSHS---LPSEASSQOVKTNGISDGKRRESPLKIDPPEDLSFN 1331
QY 481 AAAATFORASTAAGPASNVADPYRQLQYCPVSSEPSYKSG-----PALPPEGTLARSPSI 535
Db 1332 LLAVSKAQLSVQTSVPPT-PDPKRLIQLPSATQSNVNTLSSVSCMPTMP--IPARSQSQ 1388
QY 536 DSIGKDPREF-----GWRDELPREVIOMLQHOFP 564
Db 1389 ENMRSSNPFTITGLTNTNPFSDRTAAPGNPFRAKSESEATSWFSKEEPTVIS---PPF 1444
QY 565 SVQSNAAAYLQHLQFGDNKTKAEIRROG 592
Db 1445 SLOP-----LGHNKGRASSSLDG 1462
```

Search completed: July 19, 2001, 16:06:39  
Job time: 199 sec



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 19, 2001, 16:06:15 ; Search time 24.53 Seconds  
(without alignments)  
3804.070 Million cell updates/sec

Title: US-09-501-171A-4

Perfect score: 1225

Sequence: 1 MFARKPPGAPLGMVDPD.....PYSELNYESHYPASPDSDW 1225

Scoring table: OLIGO  
Gapop 60.0, Gapext 60.0

Searched: 219241 seqs, 76174552 residues

Word size: 0

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	284	23.2	1247	2 T42209	neural plakophilin
2	11	0.9	911	2 S28498	gene p120 protein
3	11	0.9	1050	2 T26395	hypothetical prote
4	10	0.8	1336	2 T02736	probable SCARECROW
5	9	0.7	153	2 T31654	hypothetical prote
6	9	0.7	184	2 S78091	endocuticular prot
7	9	0.7	243	2 S28444	comb protein - pse
8	9	0.7	301	2 T50001	hypothetical prote
9	9	0.7	415	1 A34170	acrosin (EC 3.4.21
10	9	0.7	448	2 S33926	Wilms' tumor prote
11	9	0.7	449	2 A38080	Wilms' tumor suscep
12	9	0.7	449	2 A39692	Wilms' tumor prote
13	9	0.7	494	2 A42170	zinc finger protei
14	9	0.7	497	2 T27012	hypothetical prote
15	9	0.7	497	2 JC5076	myc-associated zin
16	9	0.7	512	2 S28267	phosphorin I prec
17	9	0.7	543	2 S25128	61k protein - Auto
18	9	0.7	600	2 T14463	galactonolactone d
19	9	0.7	647	2 S06450	steroid hormone re
20	9	0.7	658	2 T04831	probable serine/th
21	9	0.7	817	2 S51342	verprolin - yeast
22	9	0.7	860	1 EAMS	elastin precursor
23	9	0.7	921	2 A33718	retinoblastoma pro
24	9	0.7	928	1 RBH9	retinoblastoma-ass
25	9	0.7	967	2 T40171	probable chromosom
26	9	0.7	979	2 T50956	hypothetical prote
27	9	0.7	1333	2 T78557	N-methyl-D-asparta
28	9	0.7	1333	2 S27224	N-methyl-D-asparta
29	9	0.7	1356	1 C45219	N-methyl-D-asparta

30	9	0.7	1420	2 T37781	probable cytoskele
31	9	0.7	1684	2 T02632	hypothetical prote
32	9	0.7	1737	2 T00209	MeG8 protein - hu
33	9	0.7	1953	2 S63244	BNI1 protein - yea
34	8	0.7	10	2 A36454	trypsin-modulating
35	8	0.7	15	2 PT0037	light harvesting c
36	8	0.7	70	2 A96650	protein F2A01.6 [1
37	8	0.7	74	2 B40513	hypothetical prote
38	8	0.7	92	2 B48831	vitelline membrane
39	8	0.7	102	2 T16978	ribosomal protein
40	8	0.7	105	2 T22564	hypothetical prote
41	8	0.7	117	2 T46322	hypothetical prote
42	8	0.7	121	2 S17718	anther-specific pr
43	8	0.7	121	2 S12245	anther-specific pr
44	8	0.7	135	2 S34815	nifN protein - Rn
45	8	0.7	141	2 A34043	hypothetical proli

#### ALIGNMENTS

RESULT 1  
T42209  
neural plakophilin related arm-repeat protein NRPAP - mouse  
N:Alternate names: plakoglobin/armadillo protein  
C:Species: Mus musculus (house mouse)  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 03-Dec-1999  
C:Accession: T42209  
R:Patfenholz, R.; Franke, W.W.  
A:Title: Identification and localization of a neurally expressed member of the plakog  
A:Reference number: Z22075; MUID:98002299  
A:Accession: T42209  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-1247 <PPF>  
A:Cross-references: EMBL:U90331; NID:92580536; PID:92580537; PIDN:AAB82409.1  
A:Experimental source: brain  
C:Genetics:  
A:Gene: NRPAP  
A>Note: predominantly, if not exclusively, expressed in neural and neuroendocrine tis

Query Match 23.2%; Score 284; DB 2; Length 1247;  
Best Local Similarity 100.0%; Pred. No. 1.5e+260;  
Matches 284; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 514 SPYKSGPALPEGTILARSPSIDSIQKDPREFGWRDPPEVIOQLQHOFPSVSNAAAY 573  
|||||  
DB 511 SPYKSGPALPEGTILARSPSIDSIQKDPREFGWRDPPEVIOQLQHOFPSVSNAAAY 570  
|||||  
QY 574 LQHLFGDNKIKAEIRROGGIQLVDLLDHRMTEVHRSACALRNLYVGKANDNKIALK 633  
|||||  
DB 571 LQHLFGDNKIKAEIRROGGIQLVDLLDHRMTEVHRSACALRNLYVGKANDNKIALK 630  
|||||  
QY 634 NCGGIPALVRLRKTTDEIRRELTVGLVNLSSCDALMPILIDALVLTAAVLIIPSGW 693  
|||||  
DB 631 NCGGIPALVRLRKTTDEIRRELTVGLVNLSSCDALMPILIDALVLTAAVLIIPSGW 690  
|||||  
QY 694 ENSPIQDDRRKIDLHSSQVLRNATGCLRNVSSAGEARRRMECDGIDLDALLVYQSLAGS 753  
|||||  
DB 691 ENSPIQDDRRKIDLHSSQVLRNATGCLRNVSSAGEARRRMECDGIDLDALLVYQSLAGS 750  
|||||  
QY 754 SEIDSKTVENCYILRNLSYRLAETSGQGHMGTDLDLGLLGE 797  
|||||  
DB 751 SEIDSKTVENCYILRNLSYRLAETSGQGHMGTDLDLGLLGE 794  
|||||  
RESULT 2  
S28498  
gene p120 protein - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 12-Mar-1993 #sequence\_revision 12-Mar-1993 #text\_change 05-Nov-1999

C:Accession: I48701; S28498  
 R:Reynolds, A.B.; Herbert, L.; Cleveland, J.L.; Berg, S.T.; Gaut, J.R.  
 Oncogene 7, 2439-2445, 1992  
 A:Title: p120, a novel substrate of protein tyrosine kinase receptors and of p60v-src,  
 A:Reference number: I48701; MUID:93096477  
 A:Accession: I48701  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-911 <RES>  
 A:Cross-references: EMBL:Z17804; NID:953544; PIDN:CAV9078.1; PID:953545  
 C:Genetics:  
 A:Gene: p120  
 C:Keywords: cytoskeleton

Query Match 0.9%; Score 11; DB 2; Length 911;  
 Best Local Similarity 100.0%; Pred. No. 0.12;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 568 SNAAYLQHLIC 578  
 |||||  
 DB 384 SNAAYLQHLIC 394

RESULT 3  
 T26395  
 hypothetical protein Y105C5B.s - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: T26395  
 R:McMurray, A.  
 submitted to the EMBL Data Library, September 1999  
 A:Reference number: Z20208  
 A:Accession: T26395  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1050 <WIL>  
 A:Cross-references: EMBL:AL110479; NID:e1542153; PIDN:CAB54367.1; CESP:Y105C5B.s  
 A:Experimental source: clone Y105C5B  
 C:Genetics:  
 A:Gene: CESP:Y105C5B.s  
 A:introns: 40/3; 105/1; 217/3; 278/1; 422/1; 482/1; 850/1; 903/1; 959/2

Query Match 0.9%; Score 11; DB 2; Length 1050;  
 Best Local Similarity 100.0%; Pred. No. 0.14;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 908 VYCAVATLRN 918  
 |||||  
 DB 828 VYCAVATLRN 838

RESULT 4  
 T02736  
 probable SCARECROW gene regulator [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 02-Feb-2001  
 C:Accession: T02736; A84692  
 R:Rounsley, S.D.; Lin, X.; Kaul, S.; Shea, T.P.; Fujii, C.Y.; Mason, T.M.; Shen, M.; Rott  
 submitted to the EMBL Data Library, August 1998  
 A:Description: Arabidopsis thaliana chromosome II BAC T914 genomic sequence.  
 A:Reference number: Z14710  
 A:Accession: T02736  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1336 <ROU>  
 A:Cross-references: EMBL:AC005315; NID:93461834; PID:93461846  
 A:Experimental source: cultivar Columbia  
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;  
 Euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
 Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
 A:Reference number: A84420; MUID:20083487  
 A:Accession: A84692  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-1336 <STO>  
 A:Cross-references: GB:AE002093; NID:93461846; PIDN:ACG3232.1; GSPDB:GN00139  
 C:Genetics:  
 A:Gene: At2g29060; T914.14  
 A:Map position: 2  
 A:introns: 694/2

Query Match 0.8%; Score 10; DB 2; Length 1336;  
 Best Local Similarity 100.0%; Pred. No. 1.5;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 810 GKRRKKRSQ 819  
 |||||  
 DB 942 GKRRKKRSQ 951

RESULT 5  
 T31654  
 hypothetical protein Y57A10A.gg - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
 C:Accession: T31654  
 R:Smeye, R.  
 submitted to the EMBL Data Library, September 1999  
 A:Reference number: Z21048  
 A:Accession: T31654  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-153 <WIL>  
 A:Cross-references: EMBL:AL117195; NID:e1549729; PIDN:CAB55037.1; CESP:Y57A10A.gg  
 A:Experimental source: clone Y57A10A  
 C:Genetics:  
 A:Gene: CESP:Y57A10A.gg

Query Match 0.7%; Score 9; DB 2; Length 153;  
 Best Local Similarity 100.0%; Pred. No. 2.1;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 218 PAPPPPP 226  
 |||||  
 DB 80 PAPPPPP 88

RESULT 6  
 S78091  
 endocuticular protein Sgabd-1 - desert locust  
 C:Species: Schistocerca gregaria (desert locust)  
 C:Date: 04-Dec-1997 #sequence\_revision 12-Dec-1997 #text\_change 11-Jan-2000  
 C:Accession: S78091  
 R:Andersen, S.  
 submitted to the Protein Sequence Database, November 1997  
 A:Description: Amino acid sequence studies on endocuticular proteins from the desert  
 A:Reference number: S78091  
 A:Accession: S78091  
 A:Molecule type: protein  
 A:Residues: 1-184 <AND>  
 A:Experimental source: strajn albino; adult; abdominal cuticle  
 C:Superfamily: cuticle protein ICPI  
 C:Keywords: blocked amino end; glycoprotein; pyroglutamic acid  
 F:1-164/Product: endocuticular protein Sgabd-1 #status experimental <MAR>  
 F:1/Modified site: pyroglutamate carboxylic acid (Gln) #status experimental  
 F:24/Binding site: carbonyl (Thr) (covalent) #status experimental  
 F:29,33/Binding site: carbohydrate (Ser) (covalent) #status experimental

Query Match 0.7%; Score 9; DB 2; Length 184;

```

Best Local Similarity 100.0%; Pred. No. 2.5;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 219 APPPPPPR 227
    |||||
Db 35 APPPPPPR 43

RESULT 7
S28444
tonb protein - Pseudomonas putida
C:Species: Pseudomonas putida
C:Date: 07-May-1993 #sequence_revision 07-May-1993 #text_change 24-Nov-1999
C:Accession: S28444
R:Bitter, W.; Tommassen, J.; Weisbeek, P.J.
Mol. Microbiol. 7, 117-130, 1993
A:Title: Identification and characterization of the exbB, exbD and tonB genes of Pseudomonas putida
A:Reference number: S28442; MUID:93172953
A:Accession: S28444
A:Molecule type: DNA
A:Residues: 1-243 <BIT>
A:Cross-references: EMBL:X70139; NID:g49192; PIDN:CAA49716.1; PID:g49195
C:Genetics:
A:Gene: tonB
C:Superfamily: tonb protein

Query Match
Best Local Similarity 100.0%; Pred. No. 3.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 216 PEPAPPPP 224
    |||||
Db 53 PEPAPPPP 61

RESULT 8
T50001
hypothetical protein F12B17.270 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Sep-2000
C:Accession: T50001
R:Bevan, M.; Bancroft, I.; Mewes, H.W.; Rudd, S.; Lemcke, K.; Mayer, K.F.X.
submitted to the Protein Sequence Database, April 2000
A:Reference number: Z25026
A:Accession: T50001
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-301 <BEV>
A:Cross-references: EMBL:AL353995; GSPDB:GN00063; ATSP:F12B17.270
A:Title: Experimental source: cultivar Columbia; BAC clone F12B17
C:Genetics:
A:Gene: ATSP:F12B17.270
A:Map position: 5
C:Superfamily: RING finger homology
F:131-182/Domain: RING finger homology <RRN>

Query Match
Best Local Similarity 100.0%; Pred. No. 3.8;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 220 PPPPPPPR 228
    |||||
Db 29 PPPPPPPR 37

RESULT 9
A34170
acrosin (EC 3.4.21.10) precursor - pig
N:Alternate names: 53K fucose-binding protein
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999

```

```

C:Accession: A34170; S08994; S02428; S04940; S16657; S02780; S10695; S12968
R:Baba, T.; Kashiwabara, S.; Watanabe, K.; Itoh, H.; Michikawa, Y.; Kimura, K.; Takad
J. Biol. Chem. 264, 11920-11927, 1989
A:Title: Activation and maturation mechanisms of boar acrosin zymogen based on the de
A:Reference number: A34170; MUID:9308595
A:Accession: A34170
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-415 <BAB>
A:Cross-references: GB:J04950; NID:g164702; PIDN:AAA31131.1; PID:g164703
R:Ceohova, D.; Toepfer-Petersen, E.; Zucker, A.; Jonakova, V.
Biol. Chem. Hoppe-Seyler 371, 317-323, 1990
A:Title: IS sperminogen a modified proacrosin? Isolation, purification, and partial c
A:Reference number: S08994; MUID:90235653
A:Accession: S08994
A:Molecule type: protein
A:Residues: 'X', 18, 'X', 20-25, 'X', 27-32, 'X', 34-38, 'X', 40-50 <CEC>
R:Toepfer-Petersen, E.; Henschen, A.
FEBS Lett. 226, 38-42, 1987
A:Title: Acrosin shows zona and fucose binding, novel properties for a serine protein
A:Reference number: S02428; MUID:88083633
A:Accession: S02428
A:Molecule type: protein
A:Residues: 17-32, 40-55 <TOB>
R:Adham, I.M.; Klemm, U.; Maier, W.M.; Hoyer-Fender, S.; Tsousidou, S.; Engel, W.
Eur. J. Biochem. 182, 563-568, 1989
A:Title: Molecular cloning of proacrosin and analysis of its expression pattern in
A:Reference number: S04940; MUID:89325301
A:Accession: S04940
A:Molecule type: mRNA
A:Residues: 1-7, 9-210, 'Q', 212-216, 'VT', 219-346, 'A', 348-388, 390-393, 'GN', 396, 'LVE', 399
A:Cross-references: EMBL:X14844
A:Note: the authors translated the codon CCG for residue 240 as Ala, GCC for residue
R:Adham, I.M.
submitted to the EMBL data library, March 1989
A:Reference number: S16657
A:Accession: S16657
A:Molecule type: mRNA
A:Residues: 1-7, 9-210, 'Q', 212-216, 'VT', 219-346, 'A', 348-388, 390-398, 'KEUL' <AD2>
A:Cross-references: EMBL:X14844; NID:g1867; PIDN:CAA32948.1; PID:g1868
A:Note: the difference at the carboxyl end is due to a frameshift error
R:Baba, T.; Michikawa, Y.; Kawakura, K.; Arai, Y.
FEBS Lett. 244, 132-136, 1989
A:Title: Activation of boar proacrosin is effected by processing at both N- and C-ter
A:Reference number: S02780; MUID:89171246
A:Accession: S02780
A:Molecule type: protein
A:Residues: 17-69 <BA2>
R:Toepfer-Petersen, E.; Steinberger, M.; von Eschenbach, C.E.; Zucker, A.
FEBS Lett. 265, 51-54, 1990
A:Title: Zona pellucida-binding of boar sperm acrosin is associated with the N-termin
A:Reference number: S10695; MUID:90306316
A:Accession: S10695
A:Molecule type: protein
A:Residues: 40-62 <TO2>
R:Toepfer-Petersen, E.; Calvete, J.; Schaefer, W.; Henschen, A.
FEBS Lett. 275, 139-142, 1990
A:Title: Complete localization of the disulfide bridges and glycosylation sites in bo
A:Reference number: S12968; MUID:91085546
A:Accession: S12968
A:Molecule type: protein
A:Residues: 17-29, 34-66, 68-91, 94-121, 123-166, 171-184, 190-207, 209-216, 219-228, 231-245;
C:Superfamily: acrosin; trypsin homology
C:Keywords: glycoprotein; hydrolase; serine proteinase; sperm
F:1-16/Domain: signal sequence #status predicted <SIG>
F:17-415/Product: acrosin #status experimental <MNT>
F:17-39/Product: acrosin light (A) chain #status experimental <LCH>
F:40-415/Product: acrosin heavy (B) chain #status experimental <HCH>
F:40-283/Domain: trypsin homology <TRY>
F:300-374/Region: proline-rich
F:119, 208/Binding site: carbohydrate (Asn) (covalent) #status experimental
F:22-153, 26-160, 71-87, 175-244, 207-223, 234-264/Disulfide bonds: #status experimental
F:86, 140, 238/Active site: His, Asp, Ser #status predicted

```

```

Query Match Similarity      0.7%: Score 9; DB 2; Length 445;
Best Local Similarity      100.0%: Pred. No. 5;
Matches      9; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

OY      218 PAPPPPPP 226
      |||||
      342 PAPPPPPP 350

RESULT 10
S33926
Wt1ms' tumor protein WT1 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 13-Jan-1995 #sequence_revision 13 #Jan-1995 #text_change 10-May-1996
C:Accession: S33926
R:Sharma, P.M.; Yang, X.; Bowman, M.; Roberts, V.; Sukumar, S.
Cancer Res. 52, 6407-6412, 1992
A:Title: Molecular cloning of rat Wt1ms' tumor complementary DNA and a study of messenger
A:Reference number: S33926; MUID:9304515
A:Accession: S33926
A:Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-448 <SHA>
A:Cross-references: EMBL:X69716
C:Genetics:
A:Gene: WT1
C:Keywords: tumor suppressor

Query Match      0.7%: Score 9; DB 2; Length 448;
Best Local Similarity      100.0%: Pred. No. 5.3;
Matches      9; Conservative      0; Mismatches      0; Indels      0; Gaps      0;

OY      218 PAPPPPPP 226
      |||||
      58 PAPPPPPP 66

RESULT 11
A38080
Wt1ms tumor susceptibility protein WT1 - human
C:Species: Homo sapiens (man)
C:Date: 31-Dec-1993 #sequence_revision 31-Dec-1993 #text_change 20-Jun-2000
C:Accession: A38080; S088273; A34673; I38504; I52811; I58315; A56411; S26286
R:Gessler, M.; Konig, A.; Bruns, G.A.
Genomics 12, 807-813, 1992
A:Title: The genomic organization and expression of the WT1 gene.
A:Reference number: A38080; MUID:92241883
A:Accession: A38080
A:Molecule type: DNA
A:Residues: 1-449 <GES1>
A:Cross-references: GB:X61631; GB:S9414; NID:937991; PIDN:CAA43819.1; PID:9825731
A:Note: It is uncertain whether Met-1 is the initiator or whether translation is initiated
A:Note: sequence extracted from NCBI backbone (NCBIN:99414, NCBIN:99422, NCBIN:99479, NC
R:Gessler, M.; Poustka, A.; Cavenee, W.; Neve, R.L.; Orkin, S.H.; Bruns, G.A.P.
Nature 343, 774-778, 1990
A:Title: Homozygous deletion in Wt1ms tumours of a zinc-finger gene identified by chromo
A:Reference number: S08273; MUID:90158822
A:Accession: S08273
A:Molecule type: mRNA
A:Residues: 1-508
A:Cross-references: EMBL:X51630; NID:937977; PIDN:CAA35956.1; PID:937978
R:Call, K.M.; Glaser, T.; Ito, C.Y.; Buckley, A.J.; Pelletier, J.; Haber, D.A.; Rose, E.
Cell 60, 509-520, 1990
A:Title: Isolation and characterization of a zinc finger polypeptide gene at the human c
A:Reference number: A34673; MUID:90150277
A:Accession: A34673
A:Molecule type: mRNA
A:Residues: 85-248, 267-364, 'F', 366-386, 'T', 388-407, 411-449 <CAL>
A:Cross-references: GB:M0393; NID:9340381; PIDN:AAA36610.1; PID:9340382

```

```

R.Haber, D.A. Sohn, R.L.; Buckler, A.J.; Pelletier, J.; Call, K.M.; Housman, D.E.
Proc. Natl. Acad. Sci. U.S.A. 88, 9618-9622, 1991
A>Title: Alternative splicing and genomic structure of the Wt1ms tumor gene Wt1.
A:Reference number: A56411; MUID:92052142
R:Phelan, S.A.; Lindberg, C.; Call, K.M.
Cell Growth Differ. 5, 677-686, 1994
A>Title: Wt1ms' tumor gene, Wt1, mRNA is down-regulated during induction of erythroid
A:Accession: 138504; MUID:94368704
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-18 <PHE>
A:Cross-references: EMBL:U06486; NID:9473563; PIDN:AA62865.1; PID:9458432
R:Pelletier, J.; Bruening, W.; Kashlan, C.E.; Meuer, S.M.; Manivel, J.C.; Striegl, J.
Cell 67, 437-447, 1991
A>Title: Germ-line mutations in the Wt1ms' tumor suppressor gene are associated with a
A:Reference number: 152811; MUID:92005721
A:Accession: 152811
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 355-365 'H', 367-377 <PEL>
A:Cross-references: GB:S61513; NID:g237599; PIDN:AA620109.1; PID:g237600
A>Note: mutant form
R:Hamilton, T.B.; Barilla, K.C.; Romaniuk, P.J.
Nucleic Acids Res. 23, 277-284, 1995
A>Title: High affinity binding sites for the Wt1ms' tumour suppressor protein Wt1.
A:Reference number: 158315; MUID:95166649
A:Accession: 158315
A>Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 'MGNHHHHHHHSGHIEGRHW', 301-364, 'F', 366-386, 'T', 388-407, 411-449 <HAN>
A:Cross-references: GB:S75264; NID:g896246; PIDN:AA63443.1; PID:g896247
A>Note: this sequence is engineered
C:Genetics:
A:Gene: GDB:WT1
A:Cross-references: GDB:120496; OMIM:194070
A:Map position: 11p13-11p13
A:Introns: 148/1; 189/1; 223/2; 249/2; 266/2; 298/3; 349/1; 379/1; 410/1
A:Notes: mRNA transcripts containing both alternatively spliced regions are the most a
C:Keywords: alternative splicing; DNA binding; kidney; tumor suppressor; zinc finger
F.1-449/Product: Wt1ms tumor susceptibility protein Wt1, splice form 1 #stratus predict
F.1-407, 411-449/Product: Wt1ms tumor susceptibility protein Wt1, splice form 4 #statu
F.1-449, 267-449/Product: Wt1ms tumor susceptibility protein Wt1, splice form 3 #statu
F.1-249, 267-407, 411-449/Product: Wt1ms tumor susceptibility protein Wt1, splice form

```



Query Match 0.7%; Score 9; DB 2; Length 449;  
 Best Local Similarity 100.0%; Pred. No. 5.3;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 PAPP PPPP 226  
 |||||  
 DB 59 PAPP PPPP 67

## RESULT 13

A42170

zinc finger protein MAZ - human (fragment)

N:Alternate names: MYC-associated zinc finger protein MAZ; zinc finger protein zF87

C:Species: Homo sapiens (man)

C&gt;Date: 03-Mar-1994 #sequence\_revision 03-Mar-1994 #text\_change 03-Jun-1996

C:Accession: A42170; A46153

R:Pyrc, J.J.; Moberg, K.H.; Hall, D.J.

Biochemistry 31, 4102-4110, 1992

A:Title: Isolation of a novel cDNA encoding a zinc-finger protein that binds to two sites

A:Reference number: A42170; MUID:92232709

A:Accession: A42170

A:Status: not compared with conceptual translation

A:Molecule type: mRNA

A:Residues: 1-494 &lt;PVR&gt;

A:Cross-references: GB:J05371

A:Note: It is uncertain whether Met-18 is the initiator or whether translation is initiated

R:Bosstone, S.A.; Asselin, C.; Patel, A.J.; Marcu, K.B.

Proc. Natl. Acad. Sci. U.S.A. 89, 7452-7456, 1992

A:Title: MAZ, a zinc finger protein, binds to c-MYC and C2 gene sequences regulating transcription

A:Reference number: A46153; MUID:92366479

A:Accession: A46153

A:Molecule type: mRNA

A:Residues: 18-417, 'L', A19-494 &lt;BOS&gt;

A:Cross-references: GB:M94046

A:Experimental source: HeLa cells

C:Keywords: DNA binding; zinc finger

F:113-125/Region: alanine-rich

F:174-183/Region: alanine-rich

F:207-230/Region: zinc finger

F:296-318/Region: zinc finger

F:324-346/Region: zinc finger

F:354-368/Region: zinc finger

F:373-405/Region: zinc finger

F:409-430/Region: zinc finger

F:452-468/Region: alanine-rich

Query Match 0.7%; Score 9; DB 2; Length 494;  
 Best Local Similarity 100.0%; Pred. No. 5.8;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 PAPP PPPP 226  
 |||||  
 DB 148 PAPP PPPP 156

## RESULT 14

T27012

hypothetical protein Y48E1B.1 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C&gt;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T27012

R:McMurray, A.

Submitted to the EMBL Data Library, March 1997

A:Reference number: Z20299

A:Accession: T27012

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-497 &lt;NTL&gt;

A:Cross-references: EMBL:Z93393; PDB: CAB07688.1; GSPDB: GNO0020; CESP: Y48E1B.1

C:Genetics:  
 A:Gene: CESP:Y48E1B.1  
 A:Map position: 2  
 A:Introns: 68/3; 277/1

Query Match 0.7%; Score 9; DB 2; Length 497;  
 Best Local Similarity 100.0%; Pred. No. 5.8;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 PAPP PPPP 226  
 |||||  
 DB 355 PAPP PPPP 363

## RESULT 15

JC5076

MYC-associated zinc-finger protein - human

N:Alternate names: MAZ protein

C:Species: Homo sapiens (man)

C&gt;Date: 31-Jan-1997 #sequence\_revision 31-Jan-1997 #text\_change 05-Nov-1999

C:Accession: JC5076

R:Tsutsui, H.; Sakatsune, O.; Itakura, K.; Yokoyama, K.K.

Biochem. Biophys. Res. Commun. 226, 801-809, 1996

A:Title: Members of the MAZ family: A novel cDNA clone for MAZ from human pancreatic

A:Reference number: JC5076; MUID:96428591

A:Accession: JC5076

A:Molecule type: mRNA

A:Residues: 1-497 &lt;TSU&gt;

A:Cross-references: DDBJ:D85131; NID:q1752741; PIDN:BAI12728.1; PID:d1013410; PID:q17

A:Experimental source: pancreatic islet

C:Comment: This protein plays a role in the control of transcriptional initiation of

and between the introns of the mouse gene for immunoglobulin M-D.

C:Keywords: phosphoprotein; zinc finger

F:146,204/480/Binding site: phosphate (Ser) (covalent) (by casein kinase II) #status

F:349/Binding site: phosphate (Tyr) (covalent) #status predicted

Search completed: July 19, 2001, 16:08:37  
 Job time: 142 sec

Query Match 0.7%; Score 9; DB 2; Length 497;  
 Best Local Similarity 100.0%; Pred. No. 5.8;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 PAPP PPPP 226  
 |||||  
 DB 157 PAPP PPPP 165

Fri Jul 20 10:09:36 2001

us-09-501-171a-4.oli.rpr

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 19, 2001, 16:08:11 ; Search time 14.7 Seconds

(without alignments)  
2854.624 Million cell updates/sec

Title: US-09-501-171A-4

Perfect score: 1225

Sequence: 1 MFARKPGADPLGAMPVPO.....PYSELNVTSHYSPSPSW 1225

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched:

Word size : 0

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : SwissProt\_39.\*

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	11	0.9	911	P120_MOUSE	P30999 mus muscula
2	11	0.9	962	ARVC_HUMAN	000192 homo sapien
3	11	0.9	969	ARVC_MOUSE	P98203 mus muscula
4	9	0.7	243	TONB_PSEPU	Q05613 pseudomonas
5	9	0.7	288	FXE3_MOUSE	Q99Y74 mus muscula
6	9	0.7	415	ACRO_PIG	P08001 sus scrofa
7	9	0.7	448	WT1_RAT	P49952 rattus norv
8	9	0.7	449	WT1_HUMAN	P19544 homo sapien
9	9	0.7	449	WT1_MOUSE	P22561 mus muscula
10	9	0.7	449	WT1_PIG	P62651 sus scrofa
11	9	0.7	470	SOX8_CHICK	P57074 gallus gall
12	9	0.7	477	MAZ_HUMAN	P56270 homo sapien
13	9	0.7	512	PER1_VOLCA	P81131 volvox cart
14	9	0.7	520	WASP_MOUSE	P70315 mus muscula
15	9	0.7	543	VP61_NPVAC	Q03209 autographa
16	9	0.7	621	VP40_HSVBC	P54817 bovine herp
17	9	0.7	647	KNRL_DROME	P13054 drosophila
18	9	0.7	817	VRL1_YEAST	P37370 saccharomyc
19	9	0.7	837	SM4G_MOUSE	Q9WH77 mus muscula
20	9	0.7	839	APB1_RAT	C03450 rattus norv
21	9	0.7	860	ELS_MOUSE	P54305 mus muscula
22	9	0.7	921	RB_MOUSE	P13405 mus muscula
23	9	0.7	928	RB_HUMAN	P06400 homo sapien
24	9	0.7	1323	NME4_MOUSE	Q00391 mus muscula
25	9	0.7	1353	BN1L_YEAST	P41832 saccharomyc
26	9	0.7	2715	TRX2_HUMAN	C9UNM6 homo sapien
27	9	0.7	10	TMOE_AEDAE	P19425 aedes aegypt
28	8	0.7	121	ASF2_HELAN	P22184 hellanthus
29	8	0.7	135	NIU1_RHOCA	Q07178 rhodobacter
30	8	0.7	141	YPRO_OMEFU	P21260 owenita fusi
31	8	0.7	142	YMB8_YEAST	Q03525 saccharomyc
32	8	0.7	144	YHP5_YEAST	P38808 saccharomyc
33	8	0.7	160	EZRI_RAT	P31977 rattus norv

34	8	0.7	174	1	THB_RAT	P17491 rattus norv
35	8	0.7	182	1	SP23_TENMO	Q27022 tenebrio mo
36	8	0.7	196	1	REMS_XENIA	Q9YGP5 xenopus lae
37	8	0.7	215	1	HANI_HUMAN	Q96004 homo sapien
38	8	0.7	223	1	TONB_XANCP	Q34261 xanthomonas
39	8	0.7	231	1	CUT2_CAEEL	P34682 caenorhabdi
40	8	0.7	261	1	GSH1_MOUSE	P31335 mus muscula
41	8	0.7	271	1	DB83_CAEEL	Q9XWV0 caenorhabdi
42	8	0.7	297	1	HX1M_CHICK	Q93367 gallus gall
43	8	0.7	299	1	NMS3_MYCTU	Q10390 mycobacteri
44	8	0.7	309	1	HXA4_CHICK	P17277 gallus gall
45	8	0.7	310	1	IF2B_SCHHO	P56329 schistosach

## ALIGNMENTS

RESULT 1  
ID P120\_MOUSE STANDARD: PRT: 911 AA.  
AC P30999;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 01-NOV-1997 (Rel. 35, Last annotation update)  
DE P120 PROTEIN.  
GN P120 OR CATNS.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=SWISS;  
RX MEDLINE=93096477; PubMed=1334250;  
RA Reynolds A.B., Herbert L., Cleveland J.L., Berg S.T., Gaut J.R.;  
RT "P120, a novel substrate of protein tyrosine kinase receptors and of  
RT p60v-src, is related to cadherin-binding factors beta-catenin,  
RT plakoglobin and armadillo."  
RL Oncogene 7:2439-2445(1992).  
RN [2]  
RP REVISIONS.  
RA Reynolds A.B.;  
RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.  
CC -!- P120 PHOSPHORYLATED BY PROTEIN-TYROSINE KINASES.  
CC -!- SIMILARITY: BELONGS TO THE BETA-CATENIN FAMILY.  
CC -!- SIMILARITY: CONTAINS 4 ARM REPEATS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
CC EMBL: Z17804; CAA79078.1; -  
DR PIR: S28498; S28498.  
DR MGD: MGI:105100; Catns.  
DR InterPro: IPR000225; -  
DR Pfam: PF00514; Armadillo\_seg. 4.  
DR PROSITE: PSS0176; ARM\_REPEAT; 3.  
KW Cytoskeleton; Structural protein; Phosphorylation; Repeat.  
FT REPEAT 398 437 ARM 1.  
FT REPEAT 441 481 ARM 2.  
FT REPEAT 647 687 ARM 3.  
FT REPEAT 694 733 ARM 4.  
SQ SEQUENCE 911 AA: 101731 MW; EELI8C6223948DD1 CRC64;

Query Match 0.9%; Score 11; DB 1; Length 911;  
Best Local Similarity 100.0%; Pred. NO. 0.057;  
Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 568 NNAAYLQHLIC 578  
 DB 384 NNAAYLQHLIC 394

RESULT 2  
 ARVC\_HUMAN STANDARD; PRT; 962 AA.  
 AC 000192;  
 DT 01-OCT-2000 (Rel. 40, Created)  
 DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE ARMADILLO REPEAT PROTEIN DELETED IN VELO-CARDIO-FACIAL SYNDROME.  
 GN ARVCF.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 NX NCBI\_TaxID=9606;  
 RA [1]  
 RP SEQUENCE FROM N.A. (SHORT AND LONG ISOFORMS).  
 RA MEDLINE=97271559; PubMed=9126485;  
 RA Strotkin H., O'Donnell H., Dasgupta R., Halford S., St Jore B.,  
 RA Puech A., Parimoo S., Morrow B., Skoultsi A., Weissman S.,  
 RA Scambler P., Kuchelapati R.;  
 RT "Identification of a new human catenin gene family member (ARVCF) from  
 RT the region deleted in velo-cardio-facial syndrome.";  
 RL Genomics 41:75-83(1997).  
 CC -1- FUNCTION: INVOLVED IN PROTEIN-PROTEIN INTERACTIONS AT ADHERENS  
 CC JUNCTIONS.  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A SHORT FORM AND A LONG FORM  
 CC (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: FOUND IN ALL THE EXAMINED TISSUES INCLUDING  
 CC HEART BRAIN, LIVER AND KIDNEY. FOUND AT LOW LEVEL IN LUNG.  
 CC -1- DISEASE: GENE DELETED IN VELO-CARDIO-FACIAL SYNDROME (VCFS) AND IS  
 CC HEMIZYGOSITY MAY PLAY A ROLE IN THE ETIOLOGY OF SOME OF THE  
 CC PHENOTYPES ASSOCIATED WITH VCFS CHARACTERIZED BY A WIDE SPECTRUM  
 CC PHENOTYPES, INCLUDING CONOTRONCAL HEART DEFECTS, CLEFT PALATE AND  
 CC FACIAL DYSMORPHOLOGY.  
 CC -1- SIMILARITY: BELONGS TO THE BETA-CATENIN FAMILY.  
 CC -1- SIMILARITY: CONTAINS 10 ARM REPEATS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: U51269; AAC51202.1; -;  
 DR HSSP: Q02248; 2BCT.  
 DR MIM: 602269; -;  
 DR MIM: 192430; -;  
 DR InterPro: IPR000225; -;  
 DR Pfam: PF00514; Armadillo\_seg. 4.  
 DR PROSITE: PS50176; ARM\_REPEAT; 3.  
 KM Repeat: Coiled coil; Alternative splicing.  
 FT DOMAIN 8 46  
 FT DOMAIN 607 623  
 FT DOMAIN 608 611  
 FT DOMAIN 348 387  
 FT REPEAT 390 429  
 FT REPEAT 433 473  
 FT REPEAT 468 508  
 FT REPEAT 526 565  
 FT REPEAT 575 622  
 FT REPEAT 646 686  
 FT REPEAT 699 738  
 FT REPEAT 739 781  
 FT REPEAT 782 826  
 FT REPEAT 826 869  
 FT VARSPLIC 1  
 MEDCINVASASILASYKEOEAFERLTRALEOERRHVALOL  
 ERAOQPGNVSIGMGSGOPLPMAMQOLVL -> MPALRL

FT SQ SEQUENCE 962 AA; 104641 MW; 74A1814A022FF2B1 CRC64;  
 (IN SHORT ISOFORM).

Query Match 0.9%; Score 11; DB 1; Length 962;  
 Best Local Similarity 100.0%; Pred. No. 0.06;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 569 NNAAYLQHLICF 579  
 DB 377 NNAAYLQHLICF 387

## RESULT 3

ARVC\_MOUSE STANDARD; PRT; 969 AA.  
 AC P98203;  
 DT 01-OCT-2000 (Rel. 40, Created)  
 DT 01-OCT-2000 (Rel. 40, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE ARMADILLO REPEAT PROTEIN DELETED IN VELO-CARDIO-FACIAL SYNDROME  
 GN ARVCF.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NX NCBI\_TaxID=10090;  
 RA [1]  
 RP SEQUENCE FROM N.A.  
 RA Starzinski-Powitz A., Kaufmann U., Urbich C., Zuplinger C.;  
 RT "The armadillo repeat protein ARVCF, a candidate for the velo cardio  
 RT facial syndrome forms complexes with cadherin.";  
 RL Submitted (JUN-1999) to the EMBL/Genbank/DBJ databases.  
 CC -1- FUNCTION: INVOLVED IN PROTEIN-PROTEIN INTERACTIONS AT ADHERENS  
 CC JUNCTIONS (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE BETA-CATENIN FAMILY.  
 CC -1- SIMILARITY: CONTAINS 10 ARM REPEATS.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AJ243418; -; NOT\_ANNOTATED\_CDS.  
 DR MGD: MGI:109620; Arvcf  
 DR PROSITE: PS50176; ARM\_REPEAT; 3.  
 KM Repeat: Coiled coil.  
 FT NON\_TER 1  
 FT DOMAIN 1 39  
 FT REPEAT 328 376  
 FT REPEAT 377 418  
 FT REPEAT 419 462  
 FT REPEAT 463 503  
 FT REPEAT 521 560  
 FT REPEAT 570 616  
 FT REPEAT 634 674  
 FT REPEAT 675 720  
 FT REPEAT 721 769  
 FT REPEAT 770 814  
 FT REPEAT 814 869  
 FT REPEAT 869 913  
 FT REPEAT 913 969  
 FT REPEAT 969 1013  
 FT REPEAT 1013 1053  
 FT REPEAT 1053 1093  
 FT REPEAT 1093 1133  
 FT REPEAT 1133 1173  
 FT REPEAT 1173 1213  
 FT REPEAT 1213 1253  
 FT REPEAT 1253 1293  
 FT REPEAT 1293 1333  
 FT REPEAT 1333 1373  
 FT REPEAT 1373 1413  
 FT REPEAT 1413 1453  
 FT REPEAT 1453 1493  
 FT REPEAT 1493 1533  
 FT REPEAT 1533 1573  
 FT REPEAT 1573 1613  
 FT REPEAT 1613 1653  
 FT REPEAT 1653 1693  
 FT REPEAT 1693 1733  
 FT REPEAT 1733 1773  
 FT REPEAT 1773 1813  
 FT REPEAT 1813 1853  
 FT REPEAT 1853 1893  
 FT REPEAT 1893 1933  
 FT REPEAT 1933 1973  
 FT REPEAT 1973 2013  
 FT REPEAT 2013 2053  
 FT REPEAT 2053 2093  
 FT REPEAT 2093 2133  
 FT REPEAT 2133 2173  
 FT REPEAT 2173 2213  
 FT REPEAT 2213 2253  
 FT REPEAT 2253 2293  
 FT REPEAT 2293 2333  
 FT REPEAT 2333 2373  
 FT REPEAT 2373 2413  
 FT REPEAT 2413 2453  
 FT REPEAT 2453 2493  
 FT REPEAT 2493 2533  
 FT REPEAT 2533 2573  
 FT REPEAT 2573 2613  
 FT REPEAT 2613 2653  
 FT REPEAT 2653 2693  
 FT REPEAT 2693 2733  
 FT REPEAT 2733 2773  
 FT REPEAT 2773 2813  
 FT REPEAT 2813 2853  
 FT REPEAT 2853 2893  
 FT REPEAT 2893 2933  
 FT REPEAT 2933 2973  
 FT REPEAT 2973 3013  
 FT REPEAT 3013 3053  
 FT REPEAT 3053 3093  
 FT REPEAT 3093 3133  
 FT REPEAT 3133 3173  
 FT REPEAT 3173 3213  
 FT REPEAT 3213 3253  
 FT REPEAT 3253 3293  
 FT REPEAT 3293 3333  
 FT REPEAT 3333 3373  
 FT REPEAT 3373 3413  
 FT REPEAT 3413 3453  
 FT REPEAT 3453 3493  
 FT REPEAT 3493 3533  
 FT REPEAT 3533 3573  
 FT REPEAT 3573 3613  
 FT REPEAT 3613 3653  
 FT REPEAT 3653 3693  
 FT REPEAT 3693 3733  
 FT REPEAT 3733 3773  
 FT REPEAT 3773 3813  
 FT REPEAT 3813 3853  
 FT REPEAT 3853 3893  
 FT REPEAT 3893 3933  
 FT REPEAT 3933 3973  
 FT REPEAT 3973 4013  
 FT REPEAT 4013 4053  
 FT REPEAT 4053 4093  
 FT REPEAT 4093 4133  
 FT REPEAT 4133 4173  
 FT REPEAT 4173 4213  
 FT REPEAT 4213 4253  
 FT REPEAT 4253 4293  
 FT REPEAT 4293 4333  
 FT REPEAT 4333 4373  
 FT REPEAT 4373 4413  
 FT REPEAT 4413 4453  
 FT REPEAT 4453 4493  
 FT REPEAT 4493 4533  
 FT REPEAT 4533 4573  
 FT REPEAT 4573 4613  
 FT REPEAT 4613 4653  
 FT REPEAT 4653 4693  
 FT REPEAT 4693 4733  
 FT REPEAT 4733 4773  
 FT REPEAT 4773 4813  
 FT REPEAT 4813 4853  
 FT REPEAT 4853 4893  
 FT REPEAT 4893 4933  
 FT REPEAT 4933 4973  
 FT REPEAT 4973 5013  
 FT REPEAT 5013 5053  
 FT REPEAT 5053 5093  
 FT REPEAT 5093 5133  
 FT REPEAT 5133 5173  
 FT REPEAT 5173 5213  
 FT REPEAT 5213 5253  
 FT REPEAT 5253 5293  
 FT REPEAT 5293 5333  
 FT REPEAT 5333 5373  
 FT REPEAT 5373 5413  
 FT REPEAT 5413 5453  
 FT REPEAT 5453 5493  
 FT REPEAT 5493 5533  
 FT REPEAT 5533 5573  
 FT REPEAT 5573 5613  
 FT REPEAT 5613 5653  
 FT REPEAT 5653 5693  
 FT REPEAT 5693 5733  
 FT REPEAT 5733 5773  
 FT REPEAT 5773 5813  
 FT REPEAT 5813 5853  
 FT REPEAT 5853 5893  
 FT REPEAT 5893 5933  
 FT REPEAT 5933 5973  
 FT REPEAT 5973 6013  
 FT REPEAT 6013 6053  
 FT REPEAT 6053 6093  
 FT REPEAT 6093 6133  
 FT REPEAT 6133 6173  
 FT REPEAT 6173 6213  
 FT REPEAT 6213 6253  
 FT REPEAT 6253 6293  
 FT REPEAT 6293 6333  
 FT REPEAT 6333 6373  
 FT REPEAT 6373 6413  
 FT REPEAT 6413 6453  
 FT REPEAT 6453 6493  
 FT REPEAT 6493 6533  
 FT REPEAT 6533 6573  
 FT REPEAT 6573 6613  
 FT REPEAT 6613 6653  
 FT REPEAT 6653 6693  
 FT REPEAT 6693 6733  
 FT REPEAT 6733 6773  
 FT REPEAT 6773 6813  
 FT REPEAT 6813 6853  
 FT REPEAT 6853 6893  
 FT REPEAT 6893 6933  
 FT REPEAT 6933 6973  
 FT REPEAT 6973 7013  
 FT REPEAT 7013 7053  
 FT REPEAT 7053 7093  
 FT REPEAT 7093 7133  
 FT REPEAT 7133 7173  
 FT REPEAT 7173 7213  
 FT REPEAT 7213 7253  
 FT REPEAT 7253 7293  
 FT REPEAT 7293 7333  
 FT REPEAT 7333 7373  
 FT REPEAT 7373 7413  
 FT REPEAT 7413 7453  
 FT REPEAT 7453 7493  
 FT REPEAT 7493 7533  
 FT REPEAT 7533 7573  
 FT REPEAT 7573 7613  
 FT REPEAT 7613 7653  
 FT REPEAT 7653 7693  
 FT REPEAT 7693 7733  
 FT REPEAT 7733 7773  
 FT REPEAT 7773 7813  
 FT REPEAT 7813 7853  
 FT REPEAT 7853 7893  
 FT REPEAT 7893 7933  
 FT REPEAT 7933 7973  
 FT REPEAT 7973 8013  
 FT REPEAT 8013 8053  
 FT REPEAT 8053 8093  
 FT REPEAT 8093 8133  
 FT REPEAT 8133 8173  
 FT REPEAT 8173 8213  
 FT REPEAT 8213 8253  
 FT REPEAT 8253 8293  
 FT REPEAT 8293 8333  
 FT REPEAT 8333 8373  
 FT REPEAT 8373 8413  
 FT REPEAT 8413 8453  
 FT REPEAT 8453 8493  
 FT REPEAT 8493 8533  
 FT REPEAT 8533 8573  
 FT REPEAT 8573 8613  
 FT REPEAT 8613 8653  
 FT REPEAT 8653 8693  
 FT REPEAT 8693 8733  
 FT REPEAT 8733 8773  
 FT REPEAT 8773 8813  
 FT REPEAT 8813 8853  
 FT REPEAT 8853 8893  
 FT REPEAT 8893 8933  
 FT REPEAT 8933 8973  
 FT REPEAT 8973 9013  
 FT REPEAT 9013 9053  
 FT REPEAT 9053 9093  
 FT REPEAT 9093 9133  
 FT REPEAT 9133 9173  
 FT REPEAT 9173 9213  
 FT REPEAT 9213 9253  
 FT REPEAT 9253 9293  
 FT REPEAT 9293 9333  
 FT REPEAT 9333 9373  
 FT REPEAT 9373 9413  
 FT REPEAT 9413 9453  
 FT REPEAT 9453 9493  
 FT REPEAT 9493 9533  
 FT REPEAT 9533 9573  
 FT REPEAT 9573 9613  
 FT REPEAT 9613 9653  
 FT REPEAT 9653 9693  
 FT REPEAT 9693 9733  
 FT REPEAT 9733 9773  
 FT REPEAT 9773 9813  
 FT REPEAT 9813 9853  
 FT REPEAT 9853 9893  
 FT REPEAT 9893 9933  
 FT REPEAT 9933 9973  
 FT REPEAT 9973 10013  
 FT REPEAT 10013 10053  
 FT REPEAT 10053 10093  
 FT REPEAT 10093 10133  
 FT REPEAT 10133 10173  
 FT REPEAT 10173 10213  
 FT REPEAT 10213 10253  
 FT REPEAT 10253 10293  
 FT REPEAT 10293 10333  
 FT REPEAT 10333 10373  
 FT REPEAT 10373 10413  
 FT REPEAT 10413 10453  
 FT REPEAT 10453 10493  
 FT REPEAT 10493 10533  
 FT REPEAT 10533 10573  
 FT REPEAT 10573 10613  
 FT REPEAT 10613 10653  
 FT REPEAT 10653 10693  
 FT REPEAT 10693 10733  
 FT REPEAT 10733 10773  
 FT REPEAT 10773 10813  
 FT REPEAT 10813 10853  
 FT REPEAT 10853 10893  
 FT REPEAT 10893 10933  
 FT REPEAT 10933 10973  
 FT REPEAT 10973 11013  
 FT REPEAT 11013 11053  
 FT REPEAT 11053 11093  
 FT REPEAT 11093 11133  
 FT REPEAT 11133 11173  
 FT REPEAT 11173 11213  
 FT REPEAT 11213 11253  
 FT REPEAT 11253 11293  
 FT REPEAT 11293 11333  
 FT REPEAT 11333 11373  
 FT REPEAT 11373 11413  
 FT REPEAT 11413 11453  
 FT REPEAT 11453 11493  
 FT REPEAT 11493 11533  
 FT REPEAT 11533 11573  
 FT REPEAT 11573 11613  
 FT REPEAT 11613 11653  
 FT REPEAT 11653 11693  
 FT REPEAT 11693 11733  
 FT REPEAT 11733 11773  
 FT REPEAT 11773 11813  
 FT REPEAT 11813 11853  
 FT REPEAT 11853 11893  
 FT REPEAT 11893 11933  
 FT REPEAT 11933 11973  
 FT REPEAT 11973 12013  
 FT REPEAT 12013 12053  
 FT REPEAT 12053 12093  
 FT REPEAT 12093 12133  
 FT REPEAT 12133 12173  
 FT REPEAT 12173 12213  
 FT REPEAT 12213 12253  
 FT REPEAT 12253 12293  
 FT REPEAT 12293 12333  
 FT REPEAT 12333 12373  
 FT REPEAT 12373 12413  
 FT REPEAT 12413 12453  
 FT REPEAT 12453 12493  
 FT REPEAT 12493 12533  
 FT REPEAT 12533 12573  
 FT REPEAT 12573 12613  
 FT REPEAT 12613 12653  
 FT REPEAT 12653 12693  
 FT REPEAT 12693 12733  
 FT REPEAT 12733 12773  
 FT REPEAT 12773 12813  
 FT REPEAT 12813 12853  
 FT REPEAT 12853 12893  
 FT REPEAT 12893 12933  
 FT REPEAT 12933 12973  
 FT REPEAT 12973 13013  
 FT REPEAT 13013 13053  
 FT REPEAT 13053 13093  
 FT REPEAT 13093 13133  
 FT REPEAT 13133 13173  
 FT REPEAT 13173 13213  
 FT REPEAT 13213 13253  
 FT REPEAT 13253 13293  
 FT REPEAT 13293 13333  
 FT REPEAT 13333 13373  
 FT REPEAT 13373 13413  
 FT REPEAT 13413 13453  
 FT REPEAT 13453 13493  
 FT REPEAT 13493 13533  
 FT REPEAT 13533 13573  
 FT REPEAT 13573 13613  
 FT REPEAT 13613 13653  
 FT REPEAT 13653 13693  
 FT REPEAT 13693 13733  
 FT REPEAT 13733 13773  
 FT REPEAT 13773 13813  
 FT REPEAT 13813 13853  
 FT REPEAT 13853 13893  
 FT REPEAT 13893 13933  
 FT REPEAT 13933 13973  
 FT REPEAT 13973 14013  
 FT REPEAT 14013 14053  
 FT REPEAT 14053 14093  
 FT REPEAT 14093 14133  
 FT REPEAT 14133 14173  
 FT REPEAT 14173 14213  
 FT REPEAT 14213 14253  
 FT REPEAT 14253 14293  
 FT REPEAT 14293 14333  
 FT REPEAT 14333 14373  
 FT REPEAT 14373 14413  
 FT REPEAT 14413 14453  
 FT REPEAT 14453 14493  
 FT REPEAT 14493 14533  
 FT REPEAT 14533 14573  
 FT REPEAT 14573 14613  
 FT REPEAT 14613 14653  
 FT REPEAT 14653 14693  
 FT REPEAT 14693 14733  
 FT REPEAT 14733 14773  
 FT REPEAT 14773 14813  
 FT REPEAT 14813 14853  
 FT REPEAT 14853 14893  
 FT REPEAT 14893 14933  
 FT REPEAT 14933 14973  
 FT REPEAT 14973 15013  
 FT REPEAT 15013 15053  
 FT REPEAT 15053 15093  
 FT REPEAT 15093 15133  
 FT REPEAT 15133 15173  
 FT REPEAT 15173 15213  
 FT REPEAT 15213 15253  
 FT REPEAT 15253 15293  
 FT REPEAT 15293 15333  
 FT REPEAT 15333 15373  
 FT REPEAT 15373 15413  
 FT REPEAT 15413 15453  
 FT REPEAT 15453 15493  
 FT REPEAT 15493 15533  
 FT REPEAT 15533 15573  
 FT REPEAT 15573 15613  
 FT REPEAT 15613 15653  
 FT REPEAT 15653 15693  
 FT REPEAT 15693 15733  
 FT REPEAT 15733 15773  
 FT REPEAT 15773 15813  
 FT REPEAT 15813 15853  
 FT REPEAT 15853 15893  
 FT REPEAT 15893 15933  
 FT REPEAT 15933 15973  
 FT REPEAT 15973 16013  
 FT REPEAT 16013 16053  
 FT REPEAT 16053 16093  
 FT REPEAT 16093 16133  
 FT REPEAT 16133 16173  
 FT REPEAT 16173 16213  
 FT REPEAT 16213 16253  
 FT REPEAT 16253 16293  
 FT REPEAT 16293 16333  
 FT REPEAT 16333 16373  
 FT REPEAT 16373 16413  
 FT REPEAT 16413 16453  
 FT REPEAT 16453 16493  
 FT REPEAT 16493 16533  
 FT REPEAT 16533 16573  
 FT REPEAT 16573 16613  
 FT REPEAT 16613 16653  
 FT REPEAT 16653 16693  
 FT REPEAT 16693 16733  
 FT REPEAT 16733 16773  
 FT REPEAT 16773 16813  
 FT REPEAT 16813 16853  
 FT REPEAT 16853 16893  
 FT REPEAT 16893 16933  
 FT REPEAT 16933 16973  
 FT REPEAT 16973 17013  
 FT REPEAT 17013 17053  
 FT REPEAT 17053 17093  
 FT REPEAT 17093 17133  
 FT REPEAT 17133 17173  
 FT REPEAT 17173 17213  
 FT REPEAT 17213 17253  
 FT REPEAT 17253 17293  
 FT REPEAT 17293 17333  
 FT REPEAT 17333 17373  
 FT REPEAT 17373 17413  
 FT REPEAT 17413 17453  
 FT REPEAT 17453 17493  
 FT REPEAT 17493 17533  
 FT REPEAT 17533 17573  
 FT REPEAT 17573 17613  
 FT REPEAT 17613 17653  
 FT REPEAT 17653 17693  
 FT REPEAT 17693 17733  
 FT REPEAT 17733 17773  
 FT REPEAT 17773 17813  
 FT REPEAT 17813 17853  
 FT REPEAT 17853 17893  
 FT REPEAT 17893 17933  
 FT REPEAT 17933 17973  
 FT REPEAT 17973 18013  
 FT REPEAT 18013 18053  
 FT REPEAT 18053 18093  
 FT REPEAT 18093 18133  
 FT REPEAT 18133 18173  
 FT REPEAT 18173 18213  
 FT REPEAT 18213 18253  
 FT REPEAT 18253 18293  
 FT REPEAT 18293 18333  
 FT REPEAT 18333 18373  
 FT REPEAT 18373 18413  
 FT REPEAT 18413 18453  
 FT REPEAT 18453 18493  
 FT REPEAT 18493 18533  
 FT REPEAT 18533 18573  
 FT REPEAT 18573 18613  
 FT REPEAT 18613 18653  
 FT REPEAT 18653 18693  
 FT REPEAT 18693 18733  
 FT REPEAT 18733 18773  
 FT REPEAT 18773 18813  
 FT REPEAT 18813 18853  
 FT REPEAT 18853 18893  
 FT REPEAT 18893 18933  
 FT REPEAT 18933 18973  
 FT REPEAT 18973 19013  
 FT REPEAT 19013 19053  
 FT REPEAT 19053 19093  
 FT REPEAT 19093 19133  
 FT REPEAT 19133 19173  
 FT REPEAT 19173 19213  
 FT REPEAT 19213 19253  
 FT REPEAT 19253 19293  
 FT REPEAT 19293 19333  
 FT REPEAT 19333 19373  
 FT REPEAT 19373 19413  
 FT REPEAT 19413 19453  
 FT REPEAT 19453 19493  
 FT REPEAT 19493 19533  
 FT REPEAT 19533 19573  
 FT REPEAT 19573 19613  
 FT REPEAT 19613 19653  
 FT REPEAT 19653 19693  
 FT REPEAT 19693 19733  
 FT REPEAT 19733 19773  
 FT REPEAT 19773 19813  
 FT REPEAT 19813 19853  
 FT REPEAT 19853 19893  
 FT REPEAT 19893 19933  
 FT REPEAT 19933 19973  
 FT REPEAT 19973 20013  
 FT REPEAT 20013 20053  
 FT REPEAT 20053 20093  
 FT REPEAT 20093 20133  
 FT REPEAT 20133 20173  
 FT REPEAT 20173 20213  
 FT REPEAT 20213 20253  
 FT REPEAT 20253 20293  
 FT REPEAT 20293 20333  
 FT REPEAT 20333 20373  
 FT REPEAT 20373 20413  
 FT REPEAT 20413 20453  
 FT REPEAT 20453 20493  
 FT REPEAT 20493 20533  
 FT REPEAT 20533 20573  
 FT REPEAT 20573 20613  
 FT REPEAT 20613 20653  
 FT REPEAT 20653 20693  
 FT REPEAT 20693 20733  
 FT REPEAT 20733 20773  
 FT REPEAT 20773 20813  
 FT REPEAT 20813 20853  
 FT REPEAT 20853 20893  
 FT REPEAT 20893 20933  
 FT REPEAT 20933 20973  
 FT REPEAT 20973 21013  
 FT REPEAT 21013 21053  
 FT REPEAT 21053 21093  
 FT REPEAT 21093 21133  
 FT REPEAT 21133 21173  
 FT REPEAT 21173 21213  
 FT REPEAT 21213 21253  
 FT REPEAT 21253 21293  
 FT REPEAT 21293 21333  
 FT REPEAT 21333 21373  
 FT REPEAT 21373 21413  
 FT REPEAT 21413 21453  
 FT REPEAT 21453 21493  
 FT REPEAT 21493 21533  
 FT REPEAT 21533 21573  
 FT REPEAT 21573 21613  
 FT REPEAT 21613 21653  
 FT REPEAT 21653 21693  
 FT REPEAT 21693 21733  
 FT REPEAT 21733 21773  
 FT REPEAT 21773 21813  
 FT REPEAT 21813 21853  
 FT REPEAT 21853 21893  
 FT REPEAT 21893 21933  
 FT REPEAT 21933 21973  
 FT REPEAT 21973 22013  
 FT REPEAT 22

```

RESULT 4
TONB_PSEPU STANDARD; PRT; 243 AA.
ID TONB_PSEPU
AC 005613;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE TONB PROTEIN.
GN TONB.
OS Pseudomonas putida.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WCS358;
RX MEDLINE=93172953; Pubmed=8437515;
RA Bilter W., Tomassen J., Weisbeek P.J.;
RT "Identification and characterization of the exbB, exbD and tonB genes
of Pseudomonas putida WCS358: their involvement in
ferriic-pseudobactin transport."
RL Mol. Microbiol. 7:117-130(1993).
CC -1- FUNCTION: INTERACTS WITH OUTER MEMBRANE RECEPTOR PROTEINS THAT
CARRY OUT HIGH-AFFINITY BINDING AND ENERGY DEPENDENT UPTAKE INTO
THE PERIPLASMIC SPACE OF SPECIFIC SUBSTRATES. IT COULD ACT TO
TRANSDUCE ENERGY FROM THE CYTOPLASMIC MEMBRANE TO SPECIFIC ENERGY-
REQUIRING PROCESSES IN THE OUTER MEMBRANE, RESULTING IN THE
RELEASE INTO THE PERIPLASM OF LIGANDS BOUND BY THESE OUTER
MEMBRANE PROTEINS (BY SIMILARITY).
CC -1- SUBUNIT: THE ACCESSORY PROTEINS EXBBD AND EXBD SEEM TO FORM A
COMPLEX WITH TONB.
CC -1- SUBCELLULAR LOCATION: PERIPLASMIC. ANCHORED TO THE CYTOPLASMIC
MEMBRANE VIA ITS N-TERMINAL SIGNAL-LIKE SEQUENCE, SPANS THE
PERIPLASM.
CC -1- SIMILARITY: BELONGS TO THE TONB FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL: X70139; CAA49716.1; -
DR PIR: S28444; S28444.
KM Transport; Protein transport; Inner membrane; Periplasmic;
KW Transmembrane; Signal-anchor; Repeat.
FT DOMAIN 1 12 CYTOPLASMIC (POTENTIAL).
FT TRANSLEM 13 32 SIGNAL-ANCHOR (POTENTIAL).
FT DOMAIN 33 243 PERIPLASMIC (POTENTIAL).
FT DOMAIN 50 61 6 X 2 AA TANDEN REPEATS OF X-P.
SQ SEQUENCE 243 AA; 25993 MW; 6575C15A147A2774 CRC64;

Query Match 0.7%; Score 9; DB 1; Length 243;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 216 PEPAPPPP 224
Db 53 PEPAPPPP 61

RESULT 5
FXE3_MOUSE STANDARD; PRT; 288 AA.
ID FXE3_MOUSE
AC 090Y14;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE FORKHEAD BOX PROTEIN E3.

```

```

GN FOXE3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV; TISSUE=Lens;
RX MEDLINE=20119184; Pubmed=10652278;
RA Blixt A., Mahlapuu M., Alitalo M., Peltto-Huikko M., Enerback S.,
RA Carlsson P.;
RT "A forkhead gene, foxe3, is essential for lens epithelial
proliferation and closure of the lens vesicle."
RL Genes Dev. 14:245-254(2000).
CC -1- FUNCTION: TRANSCRIPTION FACTOR ESSENTIAL FOR LENS EPITHELIAL
PROLIFERATION AND CLOSURE OF THE LENS VESICLE.
CC -1- SUBCELLULAR LOCATION: NUCLEAR.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED IN THE DEVELOPING LENS FROM THE
START OF LENS PLACODE INDUCTION AND BECOMES RESTRICTED TO THE
ANTERIOR PROLIFERATING CELLS WHEN LENS FIBER DIFFERENTIATION
BEGINS.
CC -1- DISEASE: DEFECTS IN FOXE3 IS A CAUSE OF DYSGENETIC LENS (DYL); IN
MOUSE MUTANT DYL THE LENS VESICLE FAILS TO SEPARATE FROM THE
ECOTDERM, CAUSING A FUSION BETWEEN THE LENS AND THE CORNEA. LACK
OF A PROLIFERATING ANTERIOR LENS EPITHELIUM LEADS TO ABSENCE OF
SECONDARY LENS FIBERS AND A DYSPLASTIC, CATARACTIC LENS.
CC -1- SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
DR EMBL: AF142647; AAF15997.1; -
DR MGD: MGI:1353569; Foxe3.
DR InterPro: IPR001766; -
DR Pfam: PF00250; Fork_head_1.
DR PRINTS: PR00053; FORKHEAD.
DR PROSITE: PS00657; FORK_HEAD_1; 1.
DR PROSITE: PS00658; FORK_HEAD_2; 1.
DR PROSITE: PS50039; FORK_HEAD_3; 1.
KW Transcription regulation; DNA-binding; Nuclear protein.
FT DOMAIN 55 58 POLY-ARG.
FT DNA_BIND 63 154 FORK-HEAD.
FT DOMAIN 166 176 POLY-PRO.
FT DOMAIN 179 186 POLY-PRO.
FT DOMAIN 229 232 POLY-ALA.
SQ SEQUENCE 288 AA; 30541 MW; 16425926F3E466C5 CRC64;

Query Match 0.7%; Score 9; DB 1; Length 288;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 218 PAPPPEPP 226
Db 166 PAPPPEPP 174

RESULT 6
ACRO_PIG STANDARD; PRT; 415 AA.
ID ACRO_PIG
AC P08001; P08000;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE ACROSIN PRECURSOR (EC 3.4.21.10) (53 KDA FUCOSE-BINDING PROTEIN).
GN ACR.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 OX NCBI\_TaxID=9823;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=89308595; PubMed=2745422;  
 RA Bada T., Kashiwabara S.I., Watanabe K., Itoh H., Michikawa Y.,  
 RA Kimura K., Takada M., Fukamizu A., Arai Y.;  
 RT "Activation and maturation mechanisms of boar acrosin zymogen based  
 RT on the deduced primary structure";  
 RL J. Biol. Chem. 264:11920-11927(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC Tissue-Testis;  
 RX MEDLINE=89325301; PubMed=2502391;  
 RA Adam I.M., Maier W.-M., Hoyer-Fender S., Tsaousidou S., Engel W.,  
 RA Klemm U.;  
 RT "Molecular cloning of preproacrosin and analysis of its expression  
 RT pattern in spermatogenesis";  
 RL Eur. J. Biochem. 182:563-568(1989).  
 RN [3]  
 RP SEQUENCE OF 17-39.  
 RC TISSUE-Sperm;  
 RX MEDLINE=84261484; PubMed=6378631;  
 RA Fock-Nuezel R., Lottspeich F., Henschen A., Mueller-Esterl W.;  
 RT "Boar acrosin is a two-chain molecule. Isolation and primary  
 RT structure of the light chain; homology with the pro-part of other  
 RT serine proteinases";  
 RL Eur. J. Biochem. 141:441-446(1984).  
 RN [4]  
 RP SEQUENCE OF 25-91.  
 RC TISSUE-Sperm;  
 RX MEDLINE=81115822; PubMed=7007202;  
 RA Fock-Nuezel R., Lottspeich F., Henschen A., Mueller-Esterl W.,  
 RA Fritz H.;  
 RT "N-terminal amino acid sequence of boar sperm acrosin. Homology with  
 RT other serine proteinases";  
 RL Hoppe-Seyler's Z. Physiol. Chem. 361:1823-1828(1980).  
 RN [5]  
 RP SEQUENCE OF 17-32 AND 40-53.  
 RC TISSUE-Sperm;  
 RX MEDLINE=88083633; PubMed=3480243;  
 RA Toepfer-Petersen E., Henschen A.;  
 RT "Acrosin shows zona and fucose binding, novel properties for a serine  
 RT proteinase";  
 RL FEBS Lett. 226:38-42(1987).  
 RN [6]  
 RP SEQUENCE OF 17-40.  
 RX MEDLINE=90253655; PubMed=2111146;  
 RA Cechova D., Toepfer-Petersen E., Zuckner A., Jonakova V.;  
 RT "Is spermogen a modified proacrosin? Isolation, purification, and  
 RT partial characterization of low-molecular-mass boar proacrosin";  
 RL Biol. Chem. Hoppe-Seyler 371:317-323(1990).  
 RN [7]  
 RP DISULFIDE BONDS, CARBOHYDRATE-LINKAGE SITES, AND PARTIAL SEQUENCE.  
 RX MEDLINE=91083546; PubMed=2261983;  
 RA Toepfer-Petersen E., Calvete J.J., Schaefer W., Henschen A.;  
 RT "Complete localization of the disulfide bridges and glycosylation  
 RT sites in boar sperm acrosin";  
 RL FEBS Lett. 275:139-142(1990).  
 CC -1- FUNCTION: ACROSIN IS THE MAJOR PROTEASE OF MAMMALIAN SPERMATOZOA.  
 CC IT IS A SERINE PROTEASE OF TRYPsin-LIKE CLEAVAGE SPECIFICITY, IT  
 CC IS SYNTHESIZED IN A ZYMOGEN FORM, PROACROSIN AND STORED IN THE  
 CC ACROSOME.  
 CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF ARG- AND LYS-BONDS; PREFERENTIAL  
 CC CLEAVAGE ARG-XAA-> LYS-LYS-> LYS-XAA.  
 CC -1- SUBUNIT: HEAVY CHAIN (CATALYTIC) AND A LIGHT CHAIN LINKED BY TWO  
 CC DISULFIDE BONDS.  
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE  
 CC TRYPsin FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: J04950; AAA3131.1; -  
 CC EMBL: X14844; CAA32948.1; -  
 CC PIR: A34170; A34170.  
 CC PIR: S02428; S02428.  
 CC PIR: S04940; S04940.  
 CC PIR: S08994; S08994.  
 CC PIR: S12968; S12968.  
 CC MEROPS: S01.223; -  
 CC InterPro: IPR001254; -  
 CC InterPro: IPR001314; -  
 CC Pfam: PF00089; trypsin; 1.  
 CC PRINTS: PR00722; CHYMOTRYPSIN.  
 CC PROSITE: PS00134; TRYPsin\_HIS; 1.  
 CC PROSITE: PS00135; TRYPsin\_SER; 1.  
 CC KEGG: Hydrolyase; Serine proteinase; Zymogen; Sperm; signal.  
 CC FT SIGNAL 1 16  
 CC FT CHAIN 17 415 ACROSIN.  
 CC FT CHAIN 17 39 ACROSIN LIGHT CHAIN.  
 CC FT CHAIN 40 338 ACROSIN HEAVY CHAIN.  
 CC FT PROPEP 339 415 PRO-RICH.  
 CC FT DISULFID 22 152 INTERCHAIN.  
 CC FT DISULFID 26 160 INTERCHAIN.  
 CC FT DISULFID 71 87  
 CC FT DISULFID 175 244  
 CC FT DISULFID 207 223  
 CC FT DISULFID 234 264  
 CC FT CARBOHYD 19 19  
 CC FT CARBOHYD 208 208  
 CC FT ACT\_SITE 86 86  
 CC FT ACT\_SITE 140 140  
 CC FT ACT\_SITE 238 238  
 CC FT CONFLICT 8 8  
 CC FT CONFLICT 211 211  
 CC FT CONFLICT 217 218  
 CC FT CONFLICT 347 347  
 CC FT CONFLICT 389 389  
 CC FT CONFLICT 399 402  
 CC FT CONFLICT 403 415  
 CC FT CONFLICT 415 415  
 CC SQ SPONDENCE 415 AA; 45387 MW; 5AD1190087E95772 CRC64;  
 CC -----  
 CC Query Match 0.7%; Score 9; DB 1; Length 415;  
 CC Best Local Similarity 100.0%; Pred. No. 2.2;  
 CC Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC -----  
 CC QY 218 PAPPPEPP 226  
 CC | | | | | | | | | |  
 CC Db 342 PAPPPEPP 350  
 CC -----  
 CC RESULT 7  
 CC ID WT1\_RAT STANDARD; PRT; 448 AA.  
 CC AC P49952;  
 CC DT 01-OCT-1996 (Rel. 34, Created)  
 CC DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 CC DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 CC DE WILMS' TUMOR PROTEIN HOMOLOG.  
 CC GN WT1 OR WT-1.  
 CC OS Rattus norvegicus (Rat).  
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 CC OX NCBI\_TaxID=10116;  
 CC RN [1]  
 CC RP SEQUENCE FROM N.A.  
 CC RC STRAIN=SPRAGUE-DAWLEY; TISSUE=Kidney;  
 CC RX MEDLINE=93046155; PubMed=1330293;  
 CC Sharma P.M., Yang X., Bowman M., Roberts V., Sukumar S.;

```

RT      "Molecular cloning of rat Wilms' tumor complementary DNA and a study
RT of messenger RNA expression in the urogenital system and the brain.",
RL      Cancer Res. 52:6407-6412(1992).
CC      CC
CC      -1- FUNCTION: POTENTIAL ROLE IN TRANSCRIPTIONAL REGULATION. RECOGNIZES
CC      AND BINDS TO THE DNA SEQUENCE 5'-CGCCCCCGC-3'.
CC      -1- SUBCELLULAR LOCATION: NUCLEAR.
CC      -1- ALTERNATIVE PRODUCTS: TWO ALTERNATIVE SPLICING SITES EXISTS.
CC      -1- TISSUE SPECIFICITY: KIDNEY (BY SIMILARITY).
CC      -1- DEVELOPMENTAL STAGE: EXPRESSED DURING KIDNEY DEVELOPMENT.
CC      -1- SIMILARITY: BELONGS TO THE EGR FAMILY OF C2H2-TYPE ZINC-FINGER
CC      PROTEINS.
CC      -----
CC      THIS SWISS-PROT ENTRY IS COPYRIGHT. IT IS PRODUCED THROUGH A COLLABORATION
CC      BETWEEN THE SWISS INSTITUTE OF BIOINFORMATICS AND THE EMBL OUTSTATION
CC      THE EUROPEAN BIOINFORMATICS INSTITUTE. THERE ARE NO RESTRICTIONS ON ITS
CC      USE BY NON-PROFIT INSTITUTIONS AS LONG AS ITS CONTENT IS IN NO WAY
CC      MODIFIED AND THIS STATEMENT IS NOT REMOVED. USAGE BY AND FOR COMMERCIAL
CC      ENTITIES REQUIRES A LICENSE AGREEMENT (SEE http://www.isb-sib.ch/announce/),
CC      OR SEND AN EMAIL TO license@isb-sib.ch).
CC      -----
DR      EMBL, X69716; CAA49373.1; -.
DR      TRANSFAC; T02352; -.
DR      InterPro; IPR000822; -.
DR      InterPro; IPR000976; -.
DR      Pfam; PF00096; zf-C2H2; 4.
DR      PRINTS; PR00048; ZINC-FINGER.
DR      PRINTS; PR00049; WILMSTUMOR.
DR      PROSITE; PS00028; ZINC_FINGER_C2H2_1; 4.
DR      PROSITE; PS50157; ZINC_FINGER_C2H2_2; 4.
KW      Zinc-finger; Metal-binding; DNA-binding; Repeat; Nuclear protein;
FT      Transcription regulation; Alternative splicing; Anti-oncogene.
FT      DOMAIN 27 82
FT      DOMAIN 322 437
FT      ZINC_FINGERS.
FT      ZN_FING 322 346
FT      C2H2-TYPE.
FT      ZN_FING 352 376
FT      C2H2-TYPE.
FT      ZN_FING 382 404
FT      C2H2-TYPE.
FT      ZN_FING 413 437
FT      C2H2-TYPE.
FT      VARSPLIC 249 265
FT      MISSING (IN ISOFORM 2 AND ISOFORM 3).
FT      VARSPLIC 407 409
FT      MISSING (IN ISOFORM 2 AND ISOFORM 4).
SQ      SEQUENCE 448 AA; 49193 MW; 329AC9AC1FF73F76 CRC64;
QY      218 PAPPPEPPP 226
DB      58 PAPPPEPPP 66

Query Match      0.7%; Score 9; DB 1; Length 448;
Best Local Similarity 100.0%; Pred. No. 2,4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 8
WT1_HUMAN
DD      WT1_HUMAN STANDARD; PRT; 449 AA.
AC      P19544;
RC      01-FEB-1991 (Rel. 17, Created)
DT      01-AUG-1991 (Rel. 19, Last sequence update)
DT      01-OCT-2000 (Rel. 40, Last annotation update)
DE      WILMS' TUMOR PROTEIN (WT33).
GN      WT1.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_TaxID=9606;
[1]
RP      SOURCE FROM N.A.
RC      TISSUE=Fetal Kidney;
RX      MEDLINE=90158822; PubMed=2154702;
RA      Gessler M., Poustka A., Cavenee W., Neve R.L., Orkin S.H.,
RA      Burns G.A.P.;
RT      "Homologous deletion in Wilms tumours of a zinc-finger gene
RT identified by chromosome jumping.";
RL      Nature 343:774-778(1990).

```

[2] SEQUENCE FROM N.A.  
RC TISSUE-Placenta;  
RX MEDLINE=92052142; PubMed=1658787;  
RA Haber D.A., Sohn R.L., Buckler A.J., Pelletier J., Call K.M.,  
RA Housman D.E.;  
RT "Alternative splicing and genomic structure of the Wilms tumor gene  
WT1.";  
RT Proc. Natl. Acad. Sci. U.S.A. 88:9618-9622(1991).  
RN [3]  
RN  
RP SEQUENCE OF 85-449 FROM N.A.  
RX MEDLINE=90150277; PubMed=154335;  
RA Call K.M., Glaser T., Ito C.Y., Buckler A.J., Pelletier J.,  
RA Haber D.A., Rose E.A., Kral A., Yeager H., Lewis W.H., Jones C.,  
RA Housman D.E.;  
RT "Isolation and characterization of a zinc finger polypeptide gene at  
RT the human chromosome 11 Wilms' tumor locus.";  
RT Cell 60:509-520(1990).  
RN [4]  
RN  
RP IDENTIFICATION OF START CODON AND ALTERNATIVE SPLICING SITES.  
RX MEDLINE=91141522; PubMed=1671709;  
RA Buckler A.J., Pelletier J., Haber D.A., Glaser T., Housman D.E.;  
RT "Isolation, characterization, and expression of the murine Wilms'  
RT tumor gene (Wt1) during kidney development.";  
RT Mol. Cell. Biol. 11:1707-1712(1991).  
RN [5]  
RN  
RP VARIANT WT CYS-366.  
RX MEDLINE=92279213; PubMed=1317572;  
RA Little M.H., Prosser J., Conde A., Smith P.J., van Heyningen V.,  
RA Hastie N.D.;  
RT "Zinc finger point mutations within the WT1 gene in Wilms tumor  
RT patients.";  
RT Proc. Natl. Acad. Sci. U.S.A. 89:4791-4795(1992).  
RN [6]  
RN  
RP VARIANTS DDS.  
RX MEDLINE=92005721; PubMed=1655284.  
RA Pelletier J., Breuning W., Kashan C.E., Mauer S.M., Manivel J.C.,  
RA Striegel J.E., Houghton D.C., Julien C., Habib R., Fouser L.,  
RA Fine R.N., Silverman B.L., Haber D.A., Housman D.;  
RT "Germline mutations in the Wilms' tumor suppressor gene are  
RT associated with abnormal urogenital development in Denys-Drash  
RT syndrome.";  
RT Cell 67:437-447(1991).  
RN [7]  
RN  
RP VARIANTS DDS.  
RX MEDLINE=93265053; PubMed=1338906;  
RA Baird P.N., Santos A., Groves N., Jadresic L., Cowell J.K.;  
RT "Constitutional mutations in the WT1 gene in patients with  
RT Denys-Drash syndrome.";  
RT Hum. Mol. Genet. 1:301-305(1992).  
RN [8]  
RN  
RP VARIANTS DDS.  
RX MEDLINE=93271983; PubMed=8388765;  
RA Little M.H., Williamson K.A., Mannens M., Kelsey A., Gosden C.,  
RA Hastie N., van Heyningen V.;  
RT "Evidence that WT1 mutations in Denys-Drash syndrome patients may act  
RT in a dominant-negative fashion.";  
RT Hum. Mol. Genet. 2:259-264(1993).  
RN [9]  
RN  
RP VARIANT MESOTHELIOMA GLY-273.  
RX MEDLINE=94004972; PubMed=8401592;  
RA Park S., Schalling M., Bernard A., Maheswaran S., Shipley G.C.,  
RA Roberts D., Fletcher J., Shipman R., Rheinwald J., Demetri G.,  
RA Griffin J., Minden M., Housman D.E., Haber D.A.;  
RT "The Wilms' tumor gene WT1 is expressed in murine mesoderm-derived  
RT tissues and mutated in a human mesothelioma.";  
RT Nat. Genet. 4:415-420(1993).  
RN [10]  
RN  
RP VARIANTS WT SER-181 AND ALA-253.  
RX MEDLINE=97268681; PubMed=9108089;  
RA Schumacher V., Schneider S., Figue A., Wildhardt G., Harms D.,  
RA Schmidt D., Weirich A., Ludwig R., Royer-Pokora B.;  
RT "Correlation of germ-line mutations and two-hit inactivation of the

RT WT1 gene with Wilms tumors of stromal-predominant histology.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 94:3972-3977(1997).  
 [11]  
 RP VARIANTS DMS TYR-377; LEU-383 AND ASN-396.  
 RX MEDLINE=98198341; PUBMED=9529364;  
 RA Jeanpierre C., Denamur E., Henry I., Cabanis M.-O., Luce S.,  
 RA Cecille A., Elion J., Peuchmaur M., Lohrat C., Maudet P.,  
 RA Gubler M.-C., Junien C.;  
 RT "Identification of constitutional WT1 mutations, in patients with  
 RT isolated diffuse mesangial sclerosis, and analysis of  
 RT genotype/phenotype correlations by use of a computerized mutation  
 RT database.";  
 RL Am. J. Hum. Genet. 62:824-833(1998).  
 [12]  
 RP REVIEW.  
 RX MEDLINE=92207913; PubMed=1313285;  
 RA Haber D.A., Buckler A.J.;  
 RT "WT1: a novel tumor suppressor gene inactivated in Wilms' tumor";  
 RL New Biol. 4:97-106(1992).  
 [13]  
 RP REVIEW.  
 RX MEDLINE=93345769; PubMed=8393820;  
 RA Rauscher F.J. III;  
 RT "The WT1 Wilms tumor gene product: a developmentally regulated  
 RT transcription factor in the kidney that functions as a tumor  
 RT suppressor.";  
 RL FASEB J. 7:896-903(1993).  
 CC -1- FUNCTION: POTENTIAL ROLE IN TRANSCRIPTIONAL REGULATION. RECOGNIZES  
 CC AND BINDS TO THE DNA SEQUENCE 5'-CGCCCCCGC-3'.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- ALTERNATIVE PRODUCTS: TWO ALTERNATIVE SPLICING SITES EXISTS.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN THE KIDNEY AND A SUBSET OF  
 CC HEMATOPOIETIC CELLS.  
 CC -1- DISEASE: WILMS TUMOR (WT) IS AN EMBRYONAL MALIGNANCY OF THE  
 CC KIDNEY THAT AFFECTS APPROXIMATELY 1 IN 10,000 INFANTS AND YOUNG  
 CC CHILDREN. IT OCCURS BOTH IN SPORADIC AND HEREDITARY FORMS.  
 CC INACTIVATION OF WT1 IS ONE OF THE CAUSES OF WILMS TUMOR.  
 CC -1- DISEASE: PATIENTS WITH DENIS-DRASH SYNDROME (DDS), WHICH IS  
 CC CHARACTERIZED BY A TYPICAL NEPHROPATHY AND GENITAL ABNORMALITIES,  
 CC HAVE DEFECTS IN THE WT1 GENE.  
 CC -1- DISEASE: DEFECTS IN WT1 ARE ALSO A CAUSE OF DIFFUSE MESANGIAL  
 CC SCLEROSIS (DMS). A FORM A DDS.  
 CC -1- SIMILARITY: BELONGS TO THE EGR FAMILY OF C2H2-TYPE ZINC-FINGER  
 CC PROTEINS.  
 CC -1- DATABASE: NAME=WT1; NOTE=WT1 mutation database;  
 CC WWW="http://www.umd.necker.fr:2003/".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL X51630; CAA35956.1; ALT\_INIT.  
 DR EMBL M80232; AAA61299.1; JOINED.  
 DR EMBL M80217; AAA61299.1; JOINED.  
 DR EMBL M80218; AAA61299.1; JOINED.  
 DR EMBL M80219; AAA61299.1; JOINED.  
 DR EMBL M80220; AAA61299.1; JOINED.  
 DR EMBL M80221; AAA61299.1; JOINED.  
 DR EMBL M80228; AAA61299.1; JOINED.  
 DR EMBL M80229; AAA61299.1; JOINED.  
 DR EMBL M80231; AAA61299.1; JOINED.  
 DR EMBL M30393; AAA36810.1; JOINED.  
 DR PIR A34673; A34673.  
 DR PIR S08273; S08273.  
 DR TRNSPAC T00899; -.  
 DR MIM 194070; -.  
 DR MIM 256370; -.  
 DR InterPro: IPR000822; -.

DR InterPro: IPR000976; -.  
 DR Pfam: PF00096; zf-C2H2; 4.  
 DR PRINTS: PR00048; ZINC\_FINGER.  
 DR PRINTS: PR00049; WILMSTUMOR.  
 DR PROSITE: PS00028; ZINC\_FINGER\_C2H2.1; 4.  
 DR PROSITE: PS0157; ZINC\_FINGER\_C2H2.2; 4.  
 KW Zinc-finger; Metal-binding; Repeat; Nuclear protein;  
 KW Transcription regulation; Alternative splicing; Anti-oncogene;  
 KW Disease mutation.  
 FT DOMAIN 27 83 PRO-RICH.  
 FT DOMAIN 323 438 ZINC\_FINGERS.  
 FT ZN\_FING 323 347 C2H2-TYPE.  
 FT ZN\_FING 353 377 C2H2-TYPE.  
 FT ZN\_FING 383 405 C2H2-TYPE.  
 FT ZN\_FING 414 438 C2H2-TYPE.  
 FT VARSPIC 250 266 MISSING (IN ISOFORM 2 AND ISOFORM 3).  
 FT VARSPIC 408 410 MISSING (IN ISOFORM 2 AND ISOFORM 4).  
 FT VARIANT 181 181 P -> S (IN WT).  
 FT VARIANT 181 181 /FTID-VAR\_007739.  
 FT VARIANT 223 223 S -> N (IN WT).  
 FT VARIANT 223 223 /FTID-VAR\_007740.  
 FT VARIANT 253 253 G -> A (IN WT).  
 FT VARIANT 253 253 /FTID-VAR\_007741.  
 FT VARIANT 273 273 S -> G (IN MESOTHELIOMA).  
 FT VARIANT 273 273 /FTID-VAR\_007742.  
 FT VARIANT 330 330 C -> Y (IN DDS).  
 FT VARIANT 360 360 /FTID-VAR\_007743.  
 FT VARIANT 360 360 C -> G (IN DDS).  
 FT VARIANT 366 366 /FTID-VAR\_007744.  
 FT VARIANT 366 366 R -> C (IN WT).  
 FT VARIANT 366 366 /FTID-VAR\_007745.  
 FT VARIANT 366 366 R -> H (IN DDS).  
 FT VARIANT 366 366 /FTID-VAR\_007746.  
 FT VARIANT 373 373 H -> Q (IN DDS).  
 Query Match 0.7%; Score 9; DB 1; Length 449;  
 Best Local Similarity 100.0%; Pred. No. 2.4;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 218 PAPPPEPP 226  
 Db 58 PAPPPEPP 66  
 RESULT 9  
 ID WT1\_MOUSE STANDARD; PRT; 449 AA.  
 AC P22561;  
 DT 01-AUG-1991 (Rel. 19, Created)  
 DT 01-AUG-1991 (Rel. 19, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE WILMS' TUMOR PROTEIN HOMOLOG.  
 GN WT1 OR WT-1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91141522; PubMed=1671709;  
 RA Buckler A.J., Pelletier J., Haber D.A., Glaser T., Housman D.E.,  
 RT "Isolation, characterization, and expression of the murine Wilms'  
 RT tumor gene (WT1) during kidney development.";  
 RL Mol. Cell. Biol. 11:1707-1712(1991).  
 CC -1- FUNCTION: POTENTIAL ROLE IN TRANSCRIPTIONAL REGULATION. RECOGNIZES  
 CC AND BINDS TO THE DNA SEQUENCE 5'-CGCCCCCGC-3'.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
 CC -1- ALTERNATIVE PRODUCTS: TWO ALTERNATIVE SPLICING SITES EXISTS.  
 CC -1- TISSUE SPECIFICITY: KIDNEY.  
 CC -1- DEVELOPMENTAL STAGE: EXPRESSED DURING KIDNEY DEVELOPMENT.  
 CC -1- SIMILARITY: BELONGS TO THE EGR FAMILY OF C2H2-TYPE ZINC-FINGER  
 CC PROTEINS.  
 CC -----



	WT1_PIG		
CC	RESULT	10	
CC	ID	WT1_PIG	STANDARD:
CC	AC	O62651;	PRT; 449 AA.
CC	DT	15-DEC-1998 (Rel. 37, Created)	
CC	DT	15-DEC-1998 (Rel. 37, Last sequence update)	
CC	DT	01-OCT-2000 (Rel. 40, Last annotation update)	
CC	DE	WILMS' TUMOR PROTEIN HOMOLOG.	
CC	GN	WT1.	
CC	OS	Sus scrofa (Pig).	
CC	OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
CC	OX	Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.	
CC	NCBI_TaxId	-9823;	
CC	(1)	SEQUENCE FROM N.A.	
CC	RP	STRAIN-LMD; RISCUE-Kidney;	
CC	RX	MEDLINE-98267201; PubMed-9602131;	
CC	RA	Tsuutani N., Oda H., Nakatsuru Y., Imai Y., Zhang S., Ueno Y.,	
CC	RA	Ishikawa T.;	
CC	RT	"CDNA cloning and developmental expression of the porcine homologue	
CC	RT	of wt1.";	
CC	RL	Gene 211:215-220(1998).	
CC	-I-	FUNCTION: POTENTIAL ROLE IN TRANSCRIPTIONAL REGULATION. RECOGNIZES	
CC	-I-	AND BINDS TO THE DNA SEQUENCE 5'-CCCCCCC-3'.	
CC	-I-	SUBCELLULAR LOCATION: NUCLEAR.	
CC	-I-	ALTERNATIVE PRODUCTS: TWO ALTERNATIVE SPICE SITES EXISTS.	
CC	-I-	DEVELOPMENTAL STAGE: EXPRESSED DURING KIDNEY DEVELOPMENT.	
CC	-I-	SIMILARITY: BELONGS TO THE EGR FAMILY OF C2H2-TYPE ZINC-FINGER	
CC		PROTEINS.	
CC		-----	
CC		This SWISS-PROT entry is copyright. It is produced through a collaboration -	
CC		between the Swiss Institute of Bioinformatics and the EMBL outstation -	

```

RESULT      11
SOXR_CHICK  SOXR_CHICK    STANDARD:          PRT;   470 AA.
AC           PSY074;
DT       01-OCT-2000 (Rel. 40, Created)
DT       01-OCT-2000 (Rel. 40, Last sequence update)
DE       01-OCT-2000 (Rel. 40, Last annotation update)
CN        TRANSCRIPTION FACTOR SOX-8.
OS        Gallus gallus (Chicken).
OC        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC        Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC        Gallus.
OX        NCBI_Taxid=9031;
RN         [1]
RP        SEQUENCE FROM N.A.
RX        MEDLINE=20302574; PubMed=10842083;
RA        Bell K.M., Western P.S., Sinclair A.H.;
RT        "SOX8 expression during chick embryogenesis.";
RL        Mech. Dev. 94:257-260(2000).
CC        -!- FUNCTION: MAY PLAY A ROLE IN CENTRAL NERVOUS SYSTEM, LIMB AND
CC        FACIAL DEVELOPMENT. MAY BE INVOLVED IN MALE SEX DETERMINATION.
CC        -!- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC        BINDS THE CONSENSUS MOTIF 5'-[AT][ATTCAATG]-3'.
CC        -!- TISSUE SPECIFICITY: WIDELY EXPRESSED IN THE EMBRYO.
CC        -!- SIMILARITY: CONTAINS 1 HMG BOX.
CC
CC        This SWISS-PROT entry is copyright. It is produced through a collaboration
CC        between the Swiss Institute of Bioinformatics and the EMBL outstation at
CC        the European Bioinformatics Institute. There are no restrictions on its
CC        use by non-profit institutions as long as its content is in no way
CC        modified and this statement is not removed. Usage by and for commercial
CC        entities requires a license agreement (See http://www.isdb-sib.ch/announce/
CC        or send an email to license@isb-sib.ch).

```

DR EMBL; AF228664; AAF73917.1; DNA-binding; Nuclear protein.  
 KM Transcription regulation; DNA-binding; Nuclear protein.  
 FT DNA\_BIND 109 177 HMG BOX.  
 SO SEQUENCE 470 AA; 50830 MW; 020ADD008AFA4334 CRC64;

Query Match 0.7%; Score 9; DB 1; Length 470;  
 Best Local Similarity 100.0%; Pred. No. 2.5;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 6 PGAPLGA 14  
 CC |||||

DB 54 PGAPLGA 62

RESULT 12  
 MAZ\_HUMAN STANDARD; PRT; 477 AA.  
 ID MAZ\_HUMAN STANDARD; PRT; 477 AA.  
 AC P56270; Q99443; Q15703;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 01-OCT-2000 (Rel. 40, Last annotation update)  
 DE MYC-ASSOCIATED ZINC FINGER PROTEIN (MAZ1) (PURINE-BINDING  
 DE TRANSCRIPTION FACTOR) (PUR-1) (ZF87) (ZF87).  
 GN MAZ.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92366479; PubMed=1502157;  
 RA Bossone S.A., Asselin C., Patel A.J., Marcu K.B.;  
 RT "MAZ, a zinc finger protein, binds to c-Myc and C2 gene sequences  
 RT regulating transcriptional initiation and termination.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 89:7452-7456(1992).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RX TISSUE-Carcinoma;  
 MEDLINE=92232709; PubMed=1567856;  
 RA Pyrc J.J., Moberg K.H., Hall D.J.;  
 RT "Isolation of a novel cDNA encoding a zinc-finger protein that binds  
 RT to two sites within the c-myc promoter.";  
 RL Biochemistry 31:4102-4110(1992).  
 [3]  
 RP SEQUENCE FROM N.A.  
 RX TISSUE-Pancreatic islets;  
 MEDLINE=96428591; PubMed=8831693;  
 RA Tsutsui H., Sakatsune O., Itakura K., Yokoyama K.K.;  
 RT "Members of the MAZ family: a novel cDNA clone for MAZ from human  
 RT pancreatic islet cells.";  
 RL Biochem. Biophys. Res. Commun. 226:801-809(1996).  
 [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=96224025; PubMed=8626793;  
 RA Parks C.L., Shenk T.;  
 RT "The serotonin 1a receptor gene contains a TATA-less promoter that  
 RT responds to MAZ and Sp1.";  
 RL J. Biol. Chem. 271:4417-4430(1996).  
 [5]  
 RP SEQUENCE FROM N.A.  
 RX TISSUE-Lymphoblastoma;  
 MEDLINE=9685418; PubMed=9685418;  
 RA Song J., Murakami H., Tsutsui H., Tang X., Matsumura M., Itakura K.,  
 RT Kanazawa I., Sun K., Yokoyama K.K.;  
 RT "Genomic organization and expression of a human gene for Myc-  
 RT associated zinc finger protein (MAZ)."  
 RL J. Biol. Chem. 273:20603-20614(1998).  
 CC -1- FUNCTION: MAY FUNCTION AS A TRANSCRIPTION FACTOR WITH DUAL ROLES  
 CC IN TRANSCRIPTION INITIATION AND TERMINATION. BINDS TO TWO SITES,  
 CC METAL AND METAL2, WITHIN THE C-MYC PROMOTER HAVING GREATER  
 CC AFFINITY FOR THE FORMER. ALSO BINDS TO MULTIPLE G/C-RICH SITES  
 CC WITHIN THE PROMOTER OF THE SPI FAMILY OF TRANSCRIPTION FACTORS.

CC -1- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).  
 CC -1- TISSUE SPECIFICITY: HEART, BRAIN, PLACENTA, LUNG, LIVER, SKELETAL  
 CC MUSCLE, AND PANCREAS. SEEMS NOT TO BE EXPRESSED IN KIDNEY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; M94046; -; NOT ANNOTATED CDS.  
 CC EMBL; M93339; -; NOT ANNOTATED CDS.  
 CC EMBL; D85131; BA12728.1; ALT\_INIT.  
 CC EMBL; U33819; AB04121.1; ALT\_INIT.  
 CC EMBL; AB017335; BAA33064.1; -.  
 CC MIM; 600999; -.  
 CC InterPro; IPR000822; -.  
 CC Pfam; PF00096; zf-C2H2; 6.  
 CC PRINTS; PR00048; ZINC\_FINGER.  
 CC PROSITE; PS00028; ZINC\_FINGER\_C2H2\_1; 5.  
 CC PROSITE; PS0157; ZINC\_FINGER\_C2H2\_2; 5.  
 CC Transcription regulation; Zinc-finger; Metal-binding; DNA-binding;  
 CC RNA-binding; Repeat; Nuclear protein.  
 KW DOMAIN 190 413 ZINC\_FINGERS.  
 FT DOMAIN 190 212  
 FT ZN\_FING 190 212 C2H2-TYPE.  
 FT ZN\_FING 279 301 C2H2-TYPE.  
 FT ZN\_FING 307 329 C2H2-TYPE.  
 FT ZN\_FING 337 360 C2H2-TYPE.  
 FT ZN\_FING 366 388 C2H2-TYPE.  
 FT ZN\_FING 392 413 C2H2-TYPE.  
 FT DOMAIN 96 108 POLY-ALA.  
 FT DOMAIN 133 139 POLY-PRO.  
 FT DOMAIN 157 161 POLY-ALA.  
 FT DOMAIN 245 249  
 FT DOMAIN 435 449 POLY-GLY.  
 FT DOMAIN 259 259  
 FT CONFLICT 259 259 MISSING (IN REF. 3).  
 FT CONFLICT 401 401 L -> M (IN REF. 2 AND 4).  
 FT CONFLICT 443 447 MISSING (IN REF. 3).  
 SO SEQUENCE 477 AA; 48607 MW; C0AC80F32C3C6825 CRC64;

Query Match 0.7%; Score 9; DB 1; Length 477;  
 Best Local Similarity 100.0%; Pred. No. 2.5;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 218 PAPPPEPP 226  
 CC |||||

DB 131 PAPPPEPP 139

RESULT 13  
 PERL\_VOICA STANDARD; PRT; 512 AA.  
 ID PERL\_VOICA STANDARD; PRT; 512 AA.  
 AC P8131;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last annotation update)  
 DE PERPHORIN I PRECURSOR.  
 OS Volvox carter;  
 OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;  
 OC Volvocaceae; Volvox.  
 OX NCBI\_TaxID=3067;  
 [1]  
 RP SEQUENCE FROM N.A. AND PARTIAL SEQUENCE.  
 RX MEDLINE=93209229; PubMed=8458341;  
 RA Sumpter M., Berg E., Wenzl S., Godl K.;  
 RT "How a sex pheromone might act at a concentration below 10(-16) M.";  
 RL EMO J. 12:831-836(1993).  
 CC -1- FUNCTION: MAY BE INVOLVED IN CONVERSION OF ASEXUAL MALES AND  
 CC FEMALES TO THE SEXUAL PATHWAY.  
 CC -1- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX.

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X69801; CAB56808.1; -
KW Signal: Extracellular matrix; Glycoprotein; Multigene family.
FT SIGNAL 1
FT CHAIN 18
FT DOMAIN 192 203 PERPHORIN I.
FT CARBOHYD 49 49 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 96 96 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 118 118 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 378 378 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 381 381 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 403 403 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 476 476 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 512 AA; 53878 MW; 091841C5D80F5E3 CRC64;

Query Match
Best Local Similarity 0.7%; Score 9; DB 1; Length 512;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 218 PAPPPEPP 226
Db 190 PAPPPEPP 198

RESULT 14
WASP_MOUSE
ID P70315; STANDARD; PRT; 520 AA.
AC 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE WISKOTT-ALDRICH SYNDROME PROTEIN HOMOLOG (WASP).
GN WASP OR WASP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C;
RA MEDLINE=96115600; PubMed=8666397;
RA Derry J.M.J., Wiedemann P., Blair P., Wang Y., Kerns J.A.,
RA Lemahieu V., Godfrey V.L., Wilkinson J.E., Francke U.;
RA "The mouse homolog of the Wiskott-Aldrich syndrome protein (WASP)
RT gene is highly conserved and maps near the scurfy (sf) mutation on
RT the X chromosome.";
RL Genomics 29:471-477(1995).
CC -1- FUNCTION: POSSIBLE REGULATOR OF LYMPHOCYTE AND PLATELET FUNCTION.
CC MAY BE INVOLVED IN SIGNALING PATHWAYS WITH CYTOSKELETAL FUNCTION.
CC (BY SIMILARITY).
CC -1- DOMAIN: THE WHI (WASP HOMOLOGY 1) DOMAIN MAY BIND A PRO-RICH
CC LIGAND.
CC -1- SIMILARITY: CONTAINS 1 WHI DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 WHI DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U54788; AAC52556.1; -

```

```

DR MGD: MGI:105059; Wasp.
DR InterPro: IPR000095;
DR InterPro: IPR001960;
DR Pfam: PF00786; PBD; 1.
DR Pfam: PF00568; WHI; 1.
DR PROSITE: PS50108; GBD; 1.
KW Repeat.
FT DOMAIN 41 147 WHI.
FT DOMAIN 240 259 GBD.
FT REPEAT 354 363 GRSGLPPXP MOTIF 1.
FT REPEAT 393 402 GRSGLPPXP MOTIF 2.
FT DOMAIN 162 167 POLY-PRO.
FT DOMAIN 314 321 POLY-PRO.
FT DOMAIN 324 341 POLY-GLY.
FT DOMAIN 368 373 POLY-PRO.
FT DOMAIN 376 379 POLY-PRO.
FT DOMAIN 384 390 POLY-PRO.
FT DOMAIN 397 403 POLY-PRO.
FT DOMAIN 408 424 POLY-PRO.
FT DOMAIN 503 520 ASP/GLU-RICH (ACIDIC).
SQ SEQUENCE 520 AA; 54191 MW; 9C223733C59F0C8A CRC64;

Query Match
Best Local Similarity 0.7%; Score 9; DB 1; Length 520;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 218 PAPPPEPP 226
Db 409 PAPPPEPP 417

RESULT 15
VP61_NPVAC
ID VP61_NPVAC STANDARD; PRT; 543 AA.
AC 003209;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE 61 KDA PROTEIN.
GN P61.
OS Autographa californica nuclear polyhedrosis virus (ACMPV).
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae.
OC Nucleopolydnavirus.
OX NCBI_TaxID=46015;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-E2;
RA MEDLINE=93138405; PubMed=1487149;
RA Pham D.O.-D., Sivasubramanian N.;
RA "Sequence and in vitro translational analysis of a 1629-nucleotide
RT ORF in Autographa californica nuclear polyhedrosis virus strain E2."
RL Gene 123:345-348(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-C6;
RA MEDLINE=92024079; PubMed=1926775;
RA Possee R.D., Sun T.P., Howard S.C., Ayres M.D., Hill-Perkins M.,
RA Gearing R.L.;
RA "Nucleotide sequence of the Autographa californica nuclear
RT polyhedrosis 9.4 kbp EcoRI-I and -R (polyhedrin gene) region.";
RL Virology 185:229-241(1991).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-C6;
RA MEDLINE=94303173; PubMed=8030224;
RA Ayres M.D., Howard S.C., Kuzio J., Lopez-Ferber M., Possee R.D.;
RA "The complete DNA sequence of Autographa californica nuclear
RT polyhedrosis virus.";
RL Virology 202:586-605(1994).
CC -1- FUNCTION: PROBABLY PLAYS AN IMPORTANT ROLE IN THE PERSISTENCE AND
CC SURVIVAL OF THE VIRUS. MAY BE A STRUCTURAL COMPONENT IN THE
CC ENVELOPE OF THE POLYHEDRON OR THE ENVELOPE OF THE POLYHEDRA-

```

CC DERIVED VIRUS (PDV).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Z11662; CAA77730.1; -;  
 DR EMBL: M75679; AAA46705.1; -;  
 DR EMBL: L22858; AAA6639.1; -;  
 DR PIR: S25128; S25128.  
 DR PIR: H40781; H40781.  
 DR DOMAIN 143 320 PRO-RICH.  
 FT DOMAIN 278 295 POLY-PRO.  
 FT CONFLICT 361 361 V -> A (IN REF. 1).  
 SQ SEQUENCE 543 AA; 60713 MW; 196D295D5E5FC100 CRC64;

Query Match 0.7%; Score 9; DB 1; Length 543;  
 Best Local Similarity 100.0%; Pred. No. 2.8;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 218 PAPPPIPP 226  
 111111111  
 Db 195 PAPPPIPP 203

Search completed: July 19, 2001, 16:11:46  
 Job time: 215 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 19, 2001, 16:04:20 ; Search time 27.58 Seconds

(without alignments)  
2692.690 Million cell updates/sec

Title: us-09-501-171a-4

Perfect score: 1225

Sequence: 1 MFARKPPGAAPLCAMPYDQ.....PYSLENYTSHVPSDPSMW 1225

Scoring table: OLIGO  
Gapop 60.0, Gapext 60.0

Searched: 412676 seqs, 60623988 residues

Word size: 0

Total number of hits satisfying chosen parameters: 412676

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: listing first 45 summaries

Database: A\_Geneseq\_0601.\*

- 1: /SIDS8/gcgdata/geneseq/geneseqp/AA1980.DAT.\*
- 2: /SIDS8/gcgdata/geneseq/geneseqp/AA1981.DAT.\*
- 3: /SIDS8/gcgdata/geneseq/geneseqp/AA1982.DAT.\*
- 4: /SIDS8/gcgdata/geneseq/geneseqp/AA1983.DAT.\*
- 5: /SIDS8/gcgdata/geneseq/geneseqp/AA1984.DAT.\*
- 6: /SIDS8/gcgdata/geneseq/geneseqp/AA1985.DAT.\*
- 7: /SIDS8/gcgdata/geneseq/geneseqp/AA1986.DAT.\*
- 8: /SIDS8/gcgdata/geneseq/geneseqp/AA1987.DAT.\*
- 9: /SIDS8/gcgdata/geneseq/geneseqp/AA1988.DAT.\*
- 10: /SIDS8/gcgdata/geneseq/geneseqp/AA1989.DAT.\*
- 11: /SIDS8/gcgdata/geneseq/geneseqp/AA1990.DAT.\*
- 12: /SIDS8/gcgdata/geneseq/geneseqp/AA1991.DAT.\*
- 13: /SIDS8/gcgdata/geneseq/geneseqp/AA1992.DAT.\*
- 14: /SIDS8/gcgdata/geneseq/geneseqp/AA1993.DAT.\*
- 15: /SIDS8/gcgdata/geneseq/geneseqp/AA1994.DAT.\*
- 16: /SIDS8/gcgdata/geneseq/geneseqp/AA1995.DAT.\*
- 17: /SIDS8/gcgdata/geneseq/geneseqp/AA1996.DAT.\*
- 18: /SIDS8/gcgdata/geneseq/geneseqp/AA1997.DAT.\*
- 19: /SIDS8/gcgdata/geneseq/geneseqp/AA1998.DAT.\*
- 20: /SIDS8/gcgdata/geneseq/geneseqp/AA1999.DAT.\*
- 21: /SIDS8/gcgdata/geneseq/geneseqp/AA2000.DAT.\*
- 22: /SIDS8/gcgdata/geneseq/geneseqp/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1225	100.0	1225	21	AAB07973
2	876	71.5	1084	20	AAW23900
3	868	70.9	1040	18	AAW24559
4	381	31.1	756	19	AAW60664
5	284	23.2	1247	21	AAW07974
6	26	2.1	1192	20	AAW23899
7	26	2.1	1211	18	AAW24560
8	26	2.1	1211	21	AAW2336
9	9	0.7	11	21	AAW36465
10	9	0.7	11	21	AAW36465
11	9	0.7	11	22	AAW30628

12	9	0.7	15	18	AAW39024
13	9	0.7	18	17	AAW05474
14	9	0.7	21	22	AAW30656
15	9	0.7	23	22	AAW30657
16	9	0.7	36	21	AAW53868
17	9	0.7	130	22	AAW29756
18	9	0.7	210	16	AAW85066
19	9	0.7	210	18	AAW22883
20	9	0.7	429	16	AAW85065
21	9	0.7	449	19	AAW47176
22	9	0.7	449	21	AAW98804
23	9	0.7	449	21	AAW98805
24	9	0.7	449	21	AAW80196
25	9	0.7	449	21	AAW80197
26	9	0.7	456	20	AAW17067
27	9	0.7	519	19	AAW70285
28	9	0.7	520	22	AAW49336
29	9	0.7	539	19	AAW48358
30	9	0.7	600	20	AAW82660
31	9	0.7	601	21	AAW79240
32	9	0.7	606	21	AAW79241
33	9	0.7	647	11	AAW6001
34	9	0.7	797	19	AAW69373
35	9	0.7	851	19	AAW69370
36	9	0.7	859	19	AAW69372
37	9	0.7	859	19	AAW69374
38	9	0.7	871	19	AAW69375
39	9	0.7	897	19	AAW69371
40	9	0.7	928	11	AAW05305
41	9	0.7	928	11	AAW05289
42	9	0.7	928	14	AAW36534
43	9	0.7	928	16	AAW74271
44	9	0.7	928	16	AAW71680
45	9	0.7	928	16	AAW71681

## ALIGNMENTS

RESULT 1	
ID AAB07973	standard; Protein; 1225 AA.
XX	
AC AAB07973:	
DT 14-NOV-2000	(first entry)
XX	
DE	A human neural plakophilin related armidillo protein.
XX	
KW	Human; presenilin; PS1; PS2; nerve cell growth; nerve damage; hNPRA;
KW	neural plakophilin related armidillo protein; Alzheimer's disease;
KW	Parkinson's disease; amyotrophic lateral sclerosis; ALS; stroke;
KW	multiple sclerosis; ischemia; stroke; neural paropathy; sciatic crush;
KW	motor neuron disease; peripheral neuropathy; neuropathy; diabetes;
KW	spinal cord injury; facial nerve crush.
XX	
OS	Homo sapiens.
XX	
PN	W0200047615-A2.
PD	
PD	17-AUG-2000.
XX	
PF	11-FEB-2000; 2000MO-CA00126.
XX	
PR	12-FEB-1999; 9905-0119835.
XX	
PA	(UTOR) UNIV TORONTO GOVERNING COUNCIL.
XX	
PI	St George-Hyslop PH, Fraser PE;
XX	
DR	WPI: 2000-524531/47.
XX	
DR	N-PSDB; AAAS9700.
XX	

Peptide resembling  
SH3-binding peptid  
Trypsin modulating  
Trypsin modulating  
Human colon cancer  
Rice glutaredoxin  
Wilms' tumour anti  
Wilms' tumour prot  
Wilms' tumour WT1  
Human WT1 protein  
Mouse WT1 protein  
Mouse Wilms' tumou  
Human Wilms' tumou  
Human 3-OST-4 prot  
Oryza sativa RNA p  
Murine WASP protei  
Fractionated X irr  
Caulliflower L-gala  
Human diarylsulfon  
Human diarylsulfon  
Insect receptor po  
Modified retinobla  
Modified retinobla  
Modified retinobla  
Modified retinobla  
Modified retinobla  
Cancer suppressin  
Predicted retinobl  
Retinoblastoma (RB  
Retinoblastoma tum  
Retinoblastoma p1  
Recombinant p110R

PT stimulation of nerve cell growth using human Neural Plakophilin Related  
 PT Armadillo Protein (hNPRAP) polypeptide, useful for the treatment of  
 PT diseases such as Alzheimer's, Parkinson's, and stroke -  
 XX  
 XX Claim 3; Page 24-27; 33pp; English.  
 CC  
 CC The present sequence represents human Neural Plakophilin Related  
 CC Armadillo Protein (hNPRAP) polypeptide. hNPRAP interacts with presenilins  
 CC (PS1 and PS2) and PS2). The specification describes a method  
 CC for stimulating the growth of nerve cells, comprising contacting  
 CC them with hNPRAP. The hNPRAP polypeptide and polynucleotide are useful  
 CC for treating nerve damage caused by a variety of diseases or physical  
 CC traumas, including Alzheimer's disease, Parkinson's disease, amyotrophic  
 CC lateral sclerosis (ALS), multiple sclerosis, stroke, ischemia associated  
 CC with stroke, neural paropathy, motor neuron diseases, sciatic crush,  
 CC peripheral neuropathy, neuropathy associated with diabetes, spinal cord  
 CC injuries and facial nerve crush.  
 CC  
 XX Sequence 1225 AA;  
 SQ  
 Query Match 100.0%; Score 1225; DB 21; Length 1225;  
 Best Local Similarity 100.0%; Pred. NO. 0;  
 Matches 1225; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MFARKPGAAPLGAMPVDPQPSASEKTSLSPLGINTSNGDSETEETTSALIASVKEQEL 60  
 DB 1 mfarppgaaplgampvdpqpsasektslsplgintsgdsetettsaliasvkeqel 60  
 QY 61 OPERITRELEAREQIVASQLECKIGSETGSMSSMSASAEQFQMSODGQKIDDELTTG 120  
 DB 61 qferitrelaeerqivaqqlerckigsetgsmssmsaeefqwsqdgqkidelittg 120  
 QY 121 LELVSCIRISOESGILPDYSTERGERPSILSOSALONSKPEGSFOYPAHYSNQTLAL 180  
 DB 121 lelvsclrisoesgilpdystergerpsilssosalonskpegsfoypahysnqtlal 180  
 QY 181 GETTPSQLPARQTQARATGOSFSOGTTSRAGHLAGEPAPPPPPPPPPPPPPPPPPPP 240  
 DB 181 gettpsqiparqtqaratgofsfsgttsragnhlagepapppppppppppppppppppp 240  
 QY 241 PDAPPAAPAAAALYSSSTSLPAPRGGSPPLAPOGGSPKLRGSGAPBAGATYAAPRGSSP 300  
 DB 241 pdappapaaaalysstslpapprggspplapoggspkllrgsgapbagaalyaaprgssp 300  
 QY 301 KQSPERLAKSYSTSPINIVVSAGLSPIRTSPPTVOSTTSSSPRHOLSTIGTYATLS 360  
 DB 301 kqspertlaksystspinivvsaglsplirtspptvosttssspriholstigtatls 360  
 QY 361 PTKRLVHASSEYKSHSQELVATLQRPGLAAGSRASYSQHGHLGDELALOSPEHHI 420  
 DB 361 ptkrlvhasseqkshsqelvatlqrpglaaagrasysqhghlgdelalospelhi 420  
 QY 421 DPYEDRYVYOKRPMKSTSSOSGDPRLPRAHNTGTYRTSTSPASGVDSVLAQTGSGOHGPN 480  
 DB 421 dpyedyrvyokrpmkstssosgdprlprahntgtyrtstspasgvdsvvlqdtgsgohgpn 480  
 QY 481 AAAATFORASTAAGPASNADPYRQLQCPVSESPYSKSGPALPEGLIARSPSIDSTOK 540  
 DB 481 aaaatforastaaagpasnadpyrqlqcpvesespyysksgpalpegliaarspsidsiqk 540  
 QY 541 DPREFGMRDPELPEVITQMLQHOFPVSQSNAAAYLQHLGFGNKTIAETIRROGGIQLVLDL 600  
 DB 541 dprefgmrdepelpevitqmlqhofpvsqsnaaaylqhlcfgnkikaetirrggqilvldl 600  
 QY 601 LDHMTVEVHRSAGCALRMLVYKGAANDKIKLKNCGGIPALVRLRKTTDEIEIETLVGV 660  
 DB 601 ldhmtvevhrsagcalrmlvykgandnkialkncggipalvrlrkttdeieitlvgyv 660  
 QY 661 LWNLSSCALKMPITODALAVLTNAVILPHSGWENSPLODDRKTQLHSSQVLRNATGCLR 720  
 DB 661 lwnlsscalkmpitodalavltnavilphsgwensploddrktqlhssqvlrnatgclr 720

QY 721 NVSSAGEARRRRRRECDGLTLDALLVYIOSALGSSSIDSKTVEENCVCILRNISYLAETS 780  
 DB 721 nvssagearrrrrrrecdgltdallvyiosalgsssidsktveencvcilrnisyriaets 780  
 QY 781 QCGHMTDELDLGLCGEANGKRAESSGCCGKKKKKKKSSQDDWDGVP.PDCAEPKGIOM 840  
 DB 781 qcghmtdeldlglcgeangkraessgccgkkkkkkssqddwdgvp.pdcaepkgiom 840  
 QY 841 LMPHSIVKRYLTLLSCSNPDTEGAGALONLWAGSMKWSYVIRAAVREKGLPIVEL 900  
 DB 841 lmphsivkryltllscsnpdtegaagalnlwagsmkwsyviraaarkekglpivel 900  
 QY 901 LRINDRVYCAVATALRNALDVNRKELIGKYMADLVHRLPGNNSNNTASKAMSDTV 960  
 DB 901 lrindrvycavatarnaldvnrkeligkymadlvhrlpgnnsnntaskamsdttv 960  
 QY 961 TANCTLHEVITRKMMENAKALRDAGIEKLVTGISTSKSGDKHSPKVVKAASQVLSMMQYR 1020  
 DB 961 tancctlhevitrkmmenakalrdagiekltvgistsksgdkhspkvkkaasqvlsmmqyr 1020  
 QY 1021 DLRSLLKKGWSQYHFFVASSSTIERDRQRPYSSSRTPSISPYRVSPNNRSASAPAPREM 1080  
 DB 1021 dlrsllkkgwsqyhffvassstierdrqrpysstrpsispyrvspnnrsasapaprem 1080  
 QY 1081 ISLKERKTDYECTGSNATYHAKGEHTSRKDMATQNTGISTLYNNSYGAPADIKHNOY 1140  
 DB 1081 islkerktdyectgsnatyhakgehtsrkdmatqntgistlynnsygapedikhnoy 1140  
 QY 1141 SAQPVQEPSRKDYETVQFONSTRYDESPEDQVHHRPPASEYTMHLGKSTGNYDF 1200  
 DB 1141 saqpvqepsrkydyetvqfonstrdydespedqvhhrrppaseytmhlgkstgnydf 1200  
 QY 1201 YSAPRYSEINETSHPASPSMW 1225  
 DB 1201 ysaarpyseinetshypaspsmw 1225

RESULT 2  
 AAY23900  
 ID AAY23900 standard; protein; 1084 AA.  
 XX  
 AC AAY23900;  
 DT 27-SEP-1999 (first entry)  
 DE Human resenilin binding armadillo protein GT24/hNPRAP.  
 XX  
 DE Human: presenilin 1; PS1; presenilin-binding protein; interacting domain;  
 KW presenilin allele; Alzheimer's disease; senile dementia;  
 KW psychiatric disease; schizophrenia; depression; neurological disease;  
 KW stroke; cerebral haemorrhage; p0071; armadillo protein.  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Misc-difference 115  
 FT /note= "not specified"  
 XX  
 PN MO9935501-A1.  
 PD 15-JUL-1999.  
 XX  
 PF 08-JAN-1999; 99WO-CA00018.  
 PR 09-JAN-1998; 98US-0070948.  
 XX  
 PA (UTOR ) UNIV TORONTO GOVERNING COUNCIL.  
 XX  
 XX Fraser PE, St George-Hyslop PH;  
 XX MPI; 1999-419410/35.  
 XX  
 PT Identifying substances that alter presenilin interactions, useful

PT for screening individuals for presenilin alleles associated with  
 PT Alzheimer's disease - useful for diagnosis of Alzheimer's disease  
 XX  
 XX Claim 9; Fig 1; 40pp; English.

XX The present sequence represents a human presenilin-binding protein,  
 CC termed GT24/human neuronal plakoglobin related armadillo protein (GT24/  
 CC hnpkbp). The specification describes a method for identifying substances  
 CC that alter the interaction of a presenilin with a presenilin-binding  
 CC protein. The method comprises contacting the interacting domain of a  
 CC presenilin protein to a presenilin-binding protein in the presence of  
 CC a test substance, and measuring the interaction of the presenilin and  
 CC the presenilin-binding protein. The method can be used to screen  
 CC individuals for presenilin alleles associated with Alzheimer's disease  
 CC and related disorders, such as senile dementia's, psychiatric diseases  
 CC such as schizophrenia and depression, and neurological disease, such as  
 CC stroke and cerebral haemorrhage.

XX Sequence 1084 AA;

Query Match 71.5%; Score 876; DB 20; Length 1084;  
 Best Local Similarity 99.8%; Pred. No. 0;  
 Matches 1076; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 148 PSLISQALQLNSKPEGSFYQYASTHSNOTALGETTBSQLEPARCTQARAGQSFSGT 207  
 DB 7 pslsqsalqlnskpegsfyqyasthsnotalgettsqparqatqatqsfsgt 66  
 QY 208 SRAGHAGPEPPAPPP 267  
 DB 67 sragnhaagpepp 126  
 QY 268 PLAAPGSGPTKLQKSGAPGEGATYAAPRGSSPKQSPSLAKSYSTSSPINIVSAGLS 327  
 DB 127 plaaapgsgptklqksgapegatyaaprgsspkqspslaksysstsspinivsaals 186  
 QY 328 PIRVTSPPVQSTISSPPIHOLSSITIGIYATLSPTKRLVHASEQYSKHSQELVATATQR 387  
 DB 187 pirtvtsppvqstissppiholssitigiyatlsptrklvhasseqyshskhsqelvatatqtr 246  
 QY 388 PGLSAGSASYSOHHGPELALQSPENHIDPIEDRYOKRPMSTLSQSGDPLP 447  
 DB 247 pglslaagsasysqhghgpealqlspenhidpiyedryokrpmstlsqsgdplp 306  
 QY 448 AHTGTYRTSTAPSPGVDSVPLQRTGSOHGPQMAAATFORASVYAGPASNADPYRQLQ 507  
 DB 307 ahtgtyrtstapspgvdsvplqrtgsohgpqmaaatforasvyaagpasnadpyrq 366  
 QY 508 YCPVSVEPTSKSGPALPPEGTIARSPSTDSIOKDRFEGWMPPELPEVYIOMIQHOFPSVQ 567  
 DB 367 ycpvsveptsksgpalppegtiarspsidsiokdrefegwmppeleviyomiqhbfpsvq 426  
 QY 568 SNAAYVLQHLICGDNKIKAEIRROGCIQLVLDLDRHTEVRSAGALRNLYVKAND 627  
 DB 427 snaayvlqlhlicgdnhkikaeirrogciqlvldldhrhtevrsagalnlyvkannd 486  
 QY 628 NKIALKNCGIPALVRLRKTDTLEIRELVGVLMNLSSCDLAKMPITQDALVLTNAV 687  
 DB 487 nkialkncgipalvrlrktdtleirelvgvlmnlsscdlakmpitqdalavltnavi 546  
 QY 688 ITHSGWENSPLODDRRKIQHLSQVYRNATGCLRNYSAGGEARRMRCCDGLTALVI 747  
 DB 547 ithsgwensploidrrkqhlsqvyrnatgclrnysaggearrmrccdgltalvi 606  
 QY 748 QGALSSSEIDSKTVENCYCILRNLSYRLAETSOGQHMGTDELGLGGENGKDAESG 807  
 DB 607 qgalssseidsktvencycilrnlsyrlaetsogqhmgtelglggenkdaessg 666  
 QY 808 CWGKKKKKKKSDQMDGVGPLEPDCAEPPKGIQMLMHPISIVKPYTLTLLSEGNPDTLEGAA 867  
 DB 667 cwgkkkkkkksgdmdgvvgplepdcappkqigmlmhpisivkpytllllsecnptlegaa 726

QY 868 GALONLAAGSKWKSIVYRAAVRKEKGLPIVELLRIDNDRVCAVATALRNALDVRNKE 927  
 DB 727 galqlaagskwksivyravvrkekglpivelrridndrvcavatarnaldvrnke 786  
 QY 928 LIGKYAMRDLYHRLPGGNNNSNNTASKAMSDDYTAOCCTLHEVITRNMEAKALRPAAGI 987  
 DB 787 ligkyamrdlyhrlpggnnnsnntaskamsddytavocctlhevitrmekalrdaagi 846  
 QY 988 EKLGYISKSKGDKSPYVVAASQVLNMMQYRDLRSLYKDKGMSQYHFAVASSSTIERDR 1047  
 DB 847 eklgyiskskgdkspyvvaasqvlnmmqyrdlrslykdkgmsqyhfvassstierdr 906  
 QY 1048 QRPYSSRTSPISISVRYVSPNNRSASAPSPREMIISLKERKTDYECTGSNATYHANGAEHT 1107  
 DB 907 qrpysstrtspisisyryvspnnrsasapspremiislkerktdyectgsnatyhgkgaeht 966  
 QY 1108 SRKAMTAMONTGISTLTKRNSYGAFAEDIKRNOVSAGQVVPQEPERKXYETQPPONSTRN 1167  
 DB 967 srkamtaqngtisltyrnsygaafaedikrnovsagqvpqeperskxyetqpponstrn 1026  
 QY 1168 DESFFEDQVHRRPASEYTMHLGLKSTGNTVDFYSAARPYSELNYESHYPAASDSMW 1225  
 DB 1027 desffedqvhrpaseytmhlglkstgntvdfysaarpyselnyeshypaasds 1084

RESULT 3  
 AAM24559  
 ID AAM24559 standard; Protein; 1040 AA.  
 XX  
 AC AAM24559;  
 XX  
 DT 06-FEB-1998 (first entry)  
 DE Presenilin-Interacting protein GT24.  
 DE  
 KW Presenilin-Interacting protein; human; Alzheimer's disease;  
 KW diagnosis; therapy; transgenic animal; animal model; GT24.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Domain 346..862  
 FT MISC-difference 12 /note= "presenilin-Interacting domain"  
 FT MISC-difference 40 /note= "encoded by GST"  
 FT MISC-difference 40 /note= "encoded by CCR"  
 FT MISC-difference 71 /note= "encoded by AKC"  
 FT WO9727296-A1.  
 PN 31-JUL-1997.  
 PD  
 XX  
 PF 27-JAN-1997; 97MO-CA00051.  
 XX  
 PR 02-JAN-1997; 97OS-0034590.  
 PR 26-JAN-1996; 96US-0592541.  
 PR 05-JUL-1996; 96US-0021573.  
 PR 12-JUL-1996; 96US-0021700.  
 PR 08-NOV-1996; 96US-0029895.  
 XX  
 XX (HSCR-) HSC RES & DEV LP.  
 PA (UTOR ) UNIV TORONTO GOVERNING COUNCIL.  
 XX  
 XX Fraser PE, Rommens JW, St George-Hyslop PH.  
 DR WPI, 1997-393684/36.  
 DR N-PDB; AAT79967.  
 XX  
 XX Presenilin-interacting protein genes - used to develop products for  
 PT the diagnosis, therapy and study of Alzheimer's disease and related  
 PT disorders

XX Claim 1; Page 99-101; 133pp; English.  
 CC GT24, a human protein with 'armadillo' repeats, has been identified  
 CC as a presenilin-interacting protein (PIP). A yeast two-hybrid kit  
 CC was employed to screen a human brain cDNA library for clones which  
 CC interacted with presenilin transmembrane 6-7 loop domain; mutations  
 CC in this loop domain are known to be causative of Alzheimer's  
 CC disease (AD). 9 PIP gene sequences (see AAT79966-74) including GT24  
 CC were identified. PIP nucleic acids, PIP proteins and peptides  
 CC (especially the presenilin interacting domain), antibodies to PIPs,  
 CC altered transformed with PIP nucleic acids, and transgenic animals  
 CC altered with PIP nucleic acids can be used for the diagnosis,  
 CC therapy and study of AD and related disorders. They can be used to  
 CC identify compounds which can modulate the expression of a PIP gene  
 CC or which bind to a PIP or modulate its activity.  
 XX  
 XX Sequence 1040 AA;

Query Match 70.9%; Score 868; DB 16; Length 1040;  
 Best Local Similarity 99.9%; Pred. No. 0;  
 Matches 968; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 257 STLPAPRPGSLAPAGSGPTKLRGSGAPGATYAAPRGSSPKSPSLAKSYSTSSP 316  
 DB 72 STLPAPRPGSLAPAGSGPTKLRGSGAPGATYAAPRGSSPKSPSLAKSYSTSSP 131  
 QY 317 INIVVSAAGSLPIRVTSPTVOSTISSPILHQLSSTICTVATLSPTKRLVHASRQYSKHS 376  
 DB 132 INIVVSAAGSLPIRVTSPTVOSTISSPILHQLSSTICTVATLSPTKRLVHASRQYSKHS 191  
 QY 377 OELVATATLORPGSLAASRASYSOSHGHLGPELRALOSPEHHIDPIYEDVYVOKPPMRS 436  
 DB 192 GELVATATLORPGSLAASRASYSOSHGHLGPELRALOSPEHHIDPIYEDVYVOKPPMRS 251  
 QY 437 ISOQSGDPLPRAHTGTVRTSTAPSSPGVSVPLQRTGSGHSPONAAATFORASVAGPA 496  
 DB 252 ISQSGDPLPRAHTGTVRTSTAPSSPGVSVPLQRTGSGHSPONAAATFORASVAGPA 311  
 QY 497 SNVAPRYQLQYCPVESPYSKSGPALPREGILARSPSIDSIQKDPREFGMRDELPIVY 556  
 DB 312 SNVAPRYQLQYCPVESPYSKSGPALPREGILARSPSIDSIQKDPREFGMRDELPIVY 371  
 QY 557 OMLQHOFPVSQNAAYLOHLCFGDNKIKAEIRROGIGIOLVLDLDRHMTVHNSAGAL 616  
 DB 372 QMLQHOFPVSQNAAYLOHLCFGDNKIKAEIRROGIGIOLVLDLDRHMTVHNSAGAL 431  
 QY 617 RNLVYRKANDDKIALKNCGGIPALVRLRKTDTLEIRELVTVGLMNLSSCDALKMPIIQ 676  
 DB 432 RNLVYRKANDDKIALKNCGGIPALVRLRKTDTLEIRELVTVGLMNLSSCDALKMPIIQ 491  
 QY 677 DALAVITNAVITPHSGWENSPLODDRKIOHSSQVLRNATGCLRNVSAGSEARRRMREC 736  
 DB 492 DALAVITNAVITPHSGWENSPLODDRKIOHSSQVLRNATGCLRNVSAGSEARRRMREC 551  
 QY 737 DELTALLVYIOSALGSSEISDKTEVENCVILRNLSYRLAETSOGQHMGTDELGLG 796  
 DB 552 DGLTALLVYIOSALGSSEISDKTEVENCVILRNLSYRLAETSOGQHMGTDELGLG 611  
 QY 797 EANGDAESSGCGWKKKKKKSSQDQWGVPLPDCAEPPKGIOMLHPSIYKPYITLISE 856  
 DB 612 EANGDAESSGCGWKKKKKKSSQDQWGVPLPDCAEPPKGIOMLHPSIYKPYITLISE 671  
 QY 857 GSNPTLEGAGALONLAAGSMKMSVYIRAAVREKGIPIVELLRITNDVAVCAVATL 916  
 DB 672 GSNPTLEGAGALONLAAGSMKMSVYIRAAVREKGIPIVELLRITNDVAVCAVATL 731  
 QY 917 RNALDVRNKEELIGRYAMDLVHRLPGGNNNSNTASKMSMDTVATVACCTLHEVITKME 976  
 DB 732 RNALDVRNKEELIGRYAMDLVHRLPGGNNNSNTASKMSMDTVATVACCTLHEVITKME 791  
 QY 977 NAKALRDAGGIEKLVGISKSGDKHSPPVYVKAASQVLSNMOWYRDLRSLYKKDKGMSQYHF 1036

DB 792 NAKALRDAGGIEKLVGISKSGDKHSPPVYVKAASQVLSNMOWYRDLRSLYKKDKGMSQYHF 851  
 QY 1037 VASSSTIERDRORPYSSTRPSISPVVSPNNKRSASAPSPREMLSKERKTDYECTGSN 1096  
 DB 852 VASSSTIERDRORPYSSTRPSISPVVSPNNKRSASAPSPREMLSKERKTDYECTGSN 911  
 QY 1097 ATYHGAKGEHTSKDKMTAQNIGISITLNRNSGAPREDIKHNQVSNQPPQPSKDYET 1156  
 DB 912 ATYHGAKGEHTSKDKMTAQNIGISITLNRNSGAPREDIKHNQVSNQPPQPSKDYET 971  
 QY 1157 YQFONSTRYDPSFEDQVHHRPPASEXTMHLGKSTGNVDFYSARPSELNVTESH 1216  
 DB 972 YQFONSTRYDPSFEDQVHHRPPASEXTMHLGKSTGNVDFYSARPSELNVTESH 1031  
 QY 1217 YPASPSWV 1225  
 DB 1032 YPASPSWV 1040

RESULT 4  
 AAW60664  
 ID AAW60664 standard; Protein; 756 AA.  
 XX  
 XX AAW60664;  
 XX  
 XX 26-OCT-1998 (first entry)  
 XX  
 XX Human ALARM protein.  
 XX  
 XX ALARM; adherens-junction linked arm protein; delta-catenin;  
 KW Presenilin 1; diagnosis; Alzheimer's disease; antibody; human.  
 KM  
 XX Homo sapiens.  
 OS

XX  
 XX Key Location/Qualifiers  
 FH Region 159..202  
 FT /note="arm repeat 1"  
 FT 203..243  
 FT /note="arm repeat 11"  
 FT 412..450  
 FT /note="arm repeat 11"  
 FT 459..499  
 FT /note="arm repeat 14"  
 FT Region  
 XX  
 XX W09825142-A1.  
 PN  
 XX 11-JUN-1998.  
 PD  
 XX  
 XX 02-DEC-1997; 97WO-US22093.  
 PE  
 XX  
 XX 02-DEC-1996; 96US-0031556.  
 PR  
 XX  
 XX (BGMH) BRIGHAM & WOMEN'S HOSPITAL.  
 PA  
 XX  
 XX Kosik KS, Zhou J;  
 PI  
 XX  
 XX WPI; 1998-333494/29.  
 DR  
 XX  
 XX N-PSDB; AAV38305.  
 DR

PT Human ALARM polypeptide(s) binding to presenilin 1 - useful, e.g. to  
 PT detect presenilin 1 and diagnose human diseases caused by mutant  
 PT form of protein interacting with ALARM  
 PS  
 XX Claim 4; Fig 1A-F; 62pp; English.  
 XX  
 XX This novel human protein, named ALARM or delta-catenin, is able to  
 CC bind to presenilin 1. Its amino acid sequence was deduced from a  
 CC cDNA clone (see AAV38305) isolated from a human brain cDNA library.  
 CC ALARM (named for adherens-junction linked arm protein) is expressed  
 CC almost exclusively in brain tissue. It contains 4 copies of the  
 CC arm repeat that is characteristic of catenin proteins. Overall,  
 CC ALARM and delta-catenin are 52.3% similar and 32.1% identical. The



CC invention provides isolated ALARM polypeptide, nucleic acids,  
CC vectors, host cells and specific antibodies. Presenilin 1 protein  
CC can be detected in samples (especially human cerebrospinal fluid),  
CC by contacting with ALARM (claimed). Mutation of the presenilin 1  
CC gene is associated with familial Alzheimer's disease. ALARM can be  
CC used to diagnose human diseases caused by a mutant form of a protein  
CC interacting with ALARM, by analysing fluid samples to detect the  
CC ALARM-interacting protein (claimed). ALARM can also be used to  
CC diagnose altered levels of presenilin 1, by contacting a sample with  
CC ALARM and determining binding (claimed). ALARM polypeptides and  
CC antibodies can be used to detect an ALARM-containing complex in  
CC biological samples, by contacting with polypeptide or antibody and  
CC determining whether the polypeptide/antibody binds to a sample  
CC component (claimed). The antibodies are also useful to detect ALARM  
CC polypeptides and to measure the effect of candidate compounds on  
CC expression or localisation of ALARM. They are useful to evaluate  
CC engineered cells prior to introduction in gene therapy, to inhibit  
CC abnormal ALARM activity or to generate anti-idiotypic antibodies,  
CC useful therapeutically.

XX Sequence 756 AA:

Query Match 31.1%; Score 381; DB 19; Length 756;

Best Local Similarity 99.6%; Pred. No. 0; Mismatches 3; Indels 0; Gaps 0;

Matches 681; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 444 PLPPAHTGTRTSTAPSSPGVDSVPLQRTGSGHGPONAAAATFORASAYAGPASNADPY 503  
DB 11 PLPPAHtgtrtstapsspgvdsvpqrtgsgqpnnaaatqrsayagpasnady 70  
QY 504 RQIQCPVSVPSPKSGPALPPEGTARSPSIDSIKDPREGWRDPELVYQIQHOF 563  
DB 71 RQIQCPVSVPSPKSGPALPPEGTARSPSIDSIKDPREGWRDPELVYQIQHOF 130  
QY 564 PSYOSNAAYALQHLCEFDNKKIKAEIRROGQIQLVLDLHBMTEVHRSAGALRNLYVK 623  
DB 131 PSYOSNAAYALQHLCEFDNKKIKAEIRRGQIQLVLDLHBMTEVHRSAGALRNLYVK 190  
QY 624 ANDNKKIALKNCGIPALVRLRLKRTDLEIRELVTVGNLNSCDALMKPIIDALAVLT 683  
DB 191 ANDNKKIALKNCGIPALVRLRLKRTDLEIRELVTVGNLNSCDALMKPIIDALAVLT 250  
QY 684 NATIPIHSGWENPIODDKRIQHSQVLRNATGCLRNVSAGEEKKRRMRREDCGTLTDL 743  
DB 251 NATIPIHSGWENPIODDKRIQHSQVLRNATGCLRNVSAGEEKKRRMRREDCGTLTDL 310  
QY 744 LVIYOSALGSEISDSTVENVCILRNLSYRLAETRSOGHMGCTDELDCGLGSEANGKDA 803  
DB 311 LVIYOSALGSEISDSTVENVCILRNLSYRLAETRSOGHMGCTDELDCGLGSEANGKDA 370  
QY 804 ESSGCGWKKKKKKSSODQWDVGPRLDCAEPKGIOMLHNPSTIVKPYLTLLSECSNPDTL 863  
DB 371 ESSGCGWKKKKKKSSODQWDVGPRLDCAEPKGIOMLHNPSTIVKPYLTLLSECSNPDTL 430  
QY 864 EGAAGALQNLAAAGSMKSYVIRAAVKEKGLPILVELLRLINDRVCAVATALRNALDY 923  
DB 431 EGAAGALQNLAAAGSMKSYVIRAAVKEKGLPILVELLRLINDRVCAVATALRNALDY 490  
QY 924 RNKELIGKYAMRLVRLFGNNSNNTASKAMSDDTVTAVCCTLHVIYIRKNNENAKALRD 983  
DB 491 RNKELIGKYAMRLVRLFGNNSNNTASKAMSDDTVTAVCCTLHVIYIRKNNENAKALRD 550  
QY 984 AGGIEKLVGISKSGDKHSKPVVKAASQVLSNMWQYRDLRLSLYKKDQWSQYHFVASSSTI 1043  
DB 551 AGGIEKLVGISKSGDKHSKPVVKAASQVLSNMWQYRDLRLSLYKKDQWSQYHFVASSSTI 610  
QY 1044 EBRORPYSSTPSTSPVAVSPNNSASAPASPREMISLKERKTYECTGSAATYHGA 1103  
DB 611 EBRORPYSSTPSTSPVAVSPNNSASAPASPREMISLKERKTYECTGSAATYHGA 670  
QY 1104 GEHTSRKAMTAONTGISTLYRNS 1127  
DB 1104 GEHTSRKAMTAONTGISTLYRNS 1127

DB 671 gehsrkdamtagnitistlyrns 694

RESULT 5  
AAB07974  
ID AAB07974 standard; Protein: 1247 AA.

AC AAB07974;

DE 14-NOV-2000 (first entry)

DE A murine neural plakophilin related armidillo protein.

XX Human; presenilin; PS1; PS2; nerve cell growth; nerve damage; hNPRAP;  
KW neural plakophilin related armidillo protein; Alzheimer's disease;  
KW Parkinson's disease; amyotrophic lateral sclerosis; ALS; stroke;  
KW multiple sclerosis; ischemia; stroke; neural paropathy; sciatic crush;  
KW motor neuron disease; peripheral neuropathy; neuropathy; diabetes;  
KW spinal cord injury; facial nerve crush.

OS Mus sp.

PN W0200047615-A2.

PD 17-AUG-2000.

PF 11-FEB-2000; 2000WO-CA00126.

PR 12-FEB-1999; 99US-0119835.

PA (UTOR ) UNIV TORONTO GOVERNING COUNCIL.

PI St George-Hyslop PH, Fraser PE;

DR WPI: 2000-524531/47.

XX N-PSDB; AAS59701.

PT Stimulation of nerve cell growth using human Neural Plakophilin Related

XX Armidillo Protein (hNPRAP) polypeptide, useful for the treatment of

PS diseases such as Alzheimer's, Parkinson's, and stroke -

XX Disclosure: Page 30-33; 33pp; English.

XX The present sequence represents Neural Plakophilin Related Armidillo  
CC Protein (NPRAP) polypeptide. Human NPRAP interacts with presenilins  
CC (PS1 and PS2). The specification describes a method  
CC for stimulating the growth of nerve cells, comprising contacting  
CC them with hNPRAP. The hNPRAP polypeptide and polynucleotide are useful  
CC for treating nerve damage caused by a variety of diseases or physical  
CC traumas, including Alzheimer's disease, Parkinson's disease, amyotrophic  
CC lateral sclerosis (ALS), multiple sclerosis, stroke, ischemia associated  
CC with stroke, neural paropathy, motor neuron diseases, sciatic crush,  
CC peripheral neuropathy, neuropathy associated with diabetes, spinal cord  
CC injuries and facial nerve crush.

XX Sequence 1247 AA:

Query Match 23.2%; Score 284; DB 21; Length 1247;

Best Local Similarity 100.0%; Pred. No. 3.6e-243; Mismatches 0; Indels 0; Gaps 0;

Matches 284; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 514 SPYKSGPALPPEGTARSPSIDSIQKDPREGWRDPELVYQIQHOFPSYOSNAAY 573  
DB 511 SPYKSGPALPPEGTARSPSIDSIQKDPREGWRDPELVYQIQHOFPSYOSNAAY 570  
QY 574 LOHLCFGDNKIKAEIRROGQIQLVLDLHBMTEVHRSAGALRNLYYKANDNKTALK 633  
DB 571 LOHLCFGDNKIKAEIRRGQIQLVLDLHBMTEVHRSAGALRNLYYKANDNKTALK 630  
QY 634 NCGGIPALVRLRLKRTDLEIRELVTVGNLNSCDALMKPIIDALAVLNNAVLIIPSGW 693  
DB 631 NCGGIPALVRLRLKRTDLEIRELVTVGNLNSCDALMKPIIDALAVLNNAVLIIPSGW 690

OY 694 ENSPLDDEK10LHSSOVLNRNATGCLRNVSAGFEARRRMRECDGLTDALLVYIOSALGS 753  
 DB 691 ensplqddkxiqhsqyrlrnatgclrnvsagfearrmrecdgltdallylvsalgs 750  
 OY 754 SEIDSKTVENCVCILRNLSYRLAETSOGQHWGTDLDGLGCE 797  
 DB 751 seidsktvencvcilrnlsyrlaetsgqgmgtldelgdlgce 794

RESULT 6  
 AAY23899 standard; protein; 1192 AA.  
 AC AAY23899;  
 DT 27-SEP-1999 (first entry)  
 DE Human resenilin binding armadillo protein p0071.  
 DE XX  
 KM Human; presenilin 1; PS1; presenilin-binding protein; interacting domain;  
 KM presentin allele; Alzheimer's disease; senile dementia;  
 KM psychiatric disease; schizophrenia; depression; neurological disease;  
 KM stroke; cerebral haemorrhage; p0071; armadillo protein.  
 OS Homo sapiens.  
 OS XX  
 PN MO9935501-AL.  
 PD 15-JUL-1999.  
 PF 08-JAN-1999; 99WO-CA00018.  
 PR 09-JAN-1998; 98US-0070948.  
 PA (UTOR) UNIV TORONTO GOVERNING COUNCIL.  
 PI Fraser PE, St George-Hyslop PH;  
 PI XX  
 DR WPI; 1999-419410/35.  
 DR XX  
 PT Identifying substances that alter presenilin interactions, useful  
 PT for screening individuals for presenilin alleles associated with  
 PT Alzheimer's disease - useful for diagnosis of Alzheimer's disease  
 PS  
 PS Claim 10; Fig 1; 40pp; English.  
 CC The present sequence represents a human presenilin-binding protein,  
 CC termed p0071. It is a member of the armadillo family of proteins.  
 CC The specification describes a method for identifying substances  
 CC that alter the interaction of a presenilin with a presenilin-binding  
 CC protein. The method comprises contacting the interacting domain of a  
 CC presenilin protein to a presenilin-binding protein in the presence of  
 CC a test substance, and measuring the interaction of the presenilin and  
 CC the presenilin-binding protein. The method can be used to screen  
 CC individuals for presenilin alleles associated with Alzheimer's disease  
 CC and related disorders, such as senile dementia's, psychiatric diseases  
 CC such as schizophrenia and depression, and neurological disease, such as  
 CC stroke and cerebral haemorrhage.  
 CC XX  
 SQ Sequence 1192 AA;

Query Match 2.1%; Score 26; DB 20; Length 1192;  
 Best Local Similarity 100.0%; Pred. No. 3.3e-14;  
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 913 ATALRNALDVNRKELIGKYAMRDIV 938  
 DB 893 atalrmaidvrnkeliqkyamrdiv 918

RESULT 7

AAW24560  
 ID AAW24560 standard; Protein; 1211 AA.  
 AC AAW24560;  
 DT 06-FEB-1998 (first entry)  
 DE Presenilin-interacting protein p0071.  
 DE XX  
 KM Presenilin-interacting protein; human; Alzheimer's disease;  
 KM diagnosis; therapy; transgenic animal; animal model; p0071.  
 OS Homo sapiens.  
 OS XX  
 FH Key Location/Qualifiers  
 FT Domain 509..1022  
 FT /note="Presenilin-interacting domain"  
 PN MO9727296-AL.  
 PD 31-JUL-1997.  
 PF 27-JAN-1997; 97WO-CA00051.  
 PR 02-JAN-1997; 97US-0034590.  
 PR 26-JUL-1996; 96US-0592541.  
 PR 05-JUL-1996; 96US-0021673.  
 PR 12-JUL-1996; 96US-0021700.  
 PR 08-NOV-1996; 96US-0029895.  
 PA (HSCR-) HSC RES & DEV LP.  
 PA (UTOR) UNIV TORONTO GOVERNING COUNCIL.  
 PI Fraser PE, Rommens JM, St George-Hyslop PH;  
 PI XX  
 DR WPI; 1997-393684/36.  
 DR N-PSDB; AAT79968.  
 DR XX  
 PT Presenilin-interacting protein genes - used to develop products for  
 PT the diagnosis, therapy and study of Alzheimer's disease and related  
 PT disorders  
 PS  
 PS Claim 1; Page 104-107; 133pp; English.  
 CC p0071, a human protein with 'armadillo' repeats, has been identified  
 CC as a presenilin-interacting protein (PIP). A yeast two-hybrid kit  
 CC was employed to screen a human brain cDNA library for clones which  
 CC interacted with presenilin transmembrane 6-7 loop domain; mutations  
 CC in this loop domain are known to be causative of Alzheimer's  
 CC disease (AD). 9 PIP gene sequences (see AAT79966-74) including p0071  
 CC were identified. PIP nucleic acids, PIP proteins and peptides  
 CC (especially the presenilin interacting domain), antibodies to PIPs,  
 CC cells transformed with PIP nucleic acids, and transgenic animals  
 CC altered with PIP nucleic acids can be used for the diagnosis,  
 CC therapy and study of AD and related disorders. They can be used to  
 CC identify compounds which can modulate the expression of a PIP gene  
 CC or which bind to a PIP or modulate its activity.  
 CC XX  
 SQ Sequence 1211 AA;

Query Match 2.1%; Score 26; DB 18; Length 1211;  
 Best Local Similarity 100.0%; Pred. No. 3.3e-14;  
 Matches 26; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 913 ATALRNALDVNRKELIGKYAMRDIV 938  
 DB 893 atalrmaidvrnkeliqkyamrdiv 918

RESULT 8  
 AAY92336 standard; Protein; 1211 AA.  
 ID AAY92336



DT 16-FEB-2001 (first entry)  
 XX Trypsin Modulating Coostatic Factor.  
 DE  
 XX Pesticidal: mosquito; Trypsin Modulating Coostatic Factor;  
 KW TMOF; hormone; trypsin biosynthesis; digestive enzyme.  
 XX  
 OS Aedes aegypti.  
 XX  
 PN WO200063235-A2.  
 XX  
 PD 26-OCT-2000.  
 XX  
 PF 18-APR-2000; 2000WO-US10375.  
 XX  
 PR 21-APR-1999; 99US-0296113.  
 XX  
 PA (UYFL ) UNIV FLORIDA RES FOUND INC.  
 XX  
 PI Borovsky D;  
 XX  
 DR WPI: 2000-687157/67.  
 XX  
 XX Recombinant plant cell transformed to express a polynucleotide encoding  
 PT a pesticidal agent such as trypsin modulating coostatic factor compounds  
 PR or neuropeptide F compounds, is useful for controlling plant pests -  
 XX  
 PS Example 2; Page 26; 48pp; English.  
 XX  
 CC The present sequence is mosquito Trypsin Modulating Coostatic Factor  
 CC (TMOF). TMOF is a hormone, which regulates the biosynthesis of the  
 CC digestive enzyme trypsin. TMOF is synthesized in the follicular  
 CC epithelium of the ovary 2-30 hours after a blood meal and is released  
 CC into the haemolymph, binding to the TMOF receptor on the midgut  
 CC epithelial cells, signalling the termination of trypsin biosynthesis.  
 CC TMOF was used to produce a recombinant plant cell, which is useful for  
 CC controlling agricultural pests, in particular insects, and is also useful  
 CC for inhibiting the production of digestive enzymes in a pest to control  
 CC the pest.  
 XX  
 SO Sequence 11 AA:

Query Match 0.7%; Score 9; DB 21; Length 11;  
 Best Local Similarity 100.0%; Pred. No. 0.74;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 PAPPPPPP 226  
 . | | | | | | | | | |  
 Db 3 PAPPPPPP 11

RESULT 11  
 AAB30628  
 ID AAB30628 standard; peptide: 11 AA.  
 XX  
 AC AAB30628;  
 XX  
 DT 19-MAR-2001 (first entry)  
 XX  
 DE Trypsin modulating coostatic factor (TMOF) pesticidal analogue.  
 XX  
 KW Hormone; trypsin modulating coostatic factor; TMOF; analogue; mosquito;  
 KW trypsin; trypsin-like enzyme; metabolic energy; insect pest; pesticide;  
 KW digestion; digestive enzyme; flesh fly; flea; sand fly; house fly;  
 KW dog fly; coleopteran; lepidopteran; dipteran; blood-sucking insect;  
 KW Diptera.  
 XX  
 OS Synthetic.  
 XX  
 PN WO200062792-A2.  
 XX  
 PD 26-OCT-2000.

XX 04-APR-2000; 2000WO-US08879.  
 PF  
 XX 21-APR-1999; 99US-0295924.  
 PR  
 XX (INSE-) INSECT BIOTECHNOLOGY INC.  
 PA (UYFL ) UNIV FLORIDA.  
 XX  
 PI Bennett J, Brandt A, Borovsky D;  
 XX  
 DR WPI: 2001-006951/01.  
 XX  
 PT New polypeptide is useful for preventing, reducing and eliminating  
 PT infestation of area by pests e.g. flesh flies or mosquito larvae,  
 PT optionally in combination with e.g. repellent, attractant, acaricide,  
 PT fungicide or herbicide -  
 XX  
 PS Claim 1; Page 58; 87pp; English.  
 XX  
 CC The present sequence represents an analogue of a peptide hormone named  
 CC trypsin modulating coostatic factor (TMOF). Mosquitoes regulate trypsin  
 CC and trypsin-like enzymes with this hormone to conserve metabolic  
 CC energy. Other insect pests use it for the same purpose. The peptide  
 CC functions as a pesticidal compound, which inhibits digestion in pests by  
 CC inhibiting synthesis of pests digestive enzymes. The pesticidal  
 CC compound is useful for preventing, reducing or eliminating infestation  
 CC of geographical areas by an insect population such as flesh flies, fleas,  
 CC sand flies, house flies and dog flies. The pesticidal compound is applied  
 CC to pest inhabited loci of the geographical area such as the body of  
 CC water inhabited by mosquito larvae, or insects such as coleopterans,  
 CC lepidopterans, dipterans or blood-sucking insects of order Diptera,  
 CC suborder Nematocera, family Colicidae or subfamily Culicinae,  
 CC Corethrinae, Ceratopogonidae and Simuliidae.  
 XX  
 SO Sequence 11 AA:

Query Match 0.7%; Score 9; DB 22; Length 11;  
 Best Local Similarity 100.0%; Pred. No. 0.74;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 PAPPPPPP 226  
 . | | | | | | | | | |  
 Db 3 PAPPPPPP 11

RESULT 12  
 AAM39024  
 ID AAM39024 standard; peptide: 15 AA.  
 XX  
 AC AAM39024;  
 XX  
 DT 27-MAR-1998 (first entry)  
 XX  
 DE Peptide resembling an SH3 domain binding peptide SEQ ID NO:423.  
 XX  
 KW Cortactin; SH3 domain; binding peptide; Src homology region 3;  
 KW tyrosine kinase; immune response; lymphokine; interleukin 1; Nck;  
 KW Abl; P13gamma; p53bp2; Crk; Yes; Grb2.  
 XX  
 OS Synthetic.  
 XX  
 PN WO9730074-A1.  
 XX  
 PD 21-AUG-1997.  
 XX  
 PF 14-FEB-1997; 97WO-US02298.  
 XX  
 PR 16-FEB-1996; 96US-0602999.  
 XX  
 PA (CYTO-) CYTOGEN CORP.  
 PA (UYNC-) UNIV NORTH CAROLINA.  
 XX



PT New polypeptide is useful for preventing, reducing and eliminating  
 PT infestation of area by pests e.g. flesh flies or mosquito larvae,  
 PT optionally in combination with e.g. repellent, attractant, acaricide,  
 PT fungicide or herbicide -

XX Example; Page 49; 87pp; English.

CC The present sequence represents an analogue of a peptide hormone named  
 CC trypsin modulating oostatic factor (TMOF). Mosquitoes regulate trypsin  
 CC and trypsin-like enzymes with this hormone to conserve metabolic  
 CC energy. Other insect pests use it for the same purpose. The peptide  
 CC functions as a pesticidal compound, which inhibits digestion in pests by  
 CC inhibiting synthesis of pests digestive enzymes. The pesticidal  
 CC compound is useful for preventing, reducing or eliminating infestation  
 CC of geographical areas by an insect population such as flesh flies, fleas,  
 CC sand flies, house flies and dog flies. The pesticidal compound is applied  
 CC to pest inhabited loci of the geographical area such as the body of  
 CC water inhabited by mosquito larvae, or insects such as coleopterans,  
 CC lepidopterans, dipterans or blood-sucking insects of order Diptera,  
 CC suborder Nematocera, family Colicidae or subfamily Culicinae,  
 CC Corethrinae, Ceratopogonidae and Simuliidae.

SQ Sequence 21 AA;

Query Match 0.7%; Score 9; DB 22; Length 21;  
 Best Local Similarity 100.0%; Pred. No. 1.3;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 PAPP PPPP 226  
 Db 3 PAPP PPPP 11

RESULT 15

AAB30657  
 ID AAB30657 standard; peptide: 23 AA.

XX AAB30657;

DT 19-MAR-2001 (first entry)

XX Trypsin modulating oostatic factor (TMOF) pesticidal analogue.

XX Hormone; trypsin modulating oostatic factor; TMOF; analogue; mosquito;  
 KW trypsin; trypsin-like enzyme; metabolic energy; insect pest; pesticide;  
 KW digestion; digestive enzyme; flesh fly; flea; sand fly; house fly;  
 KW dog fly; coleopteran; lepidopteran; dipteran; blood-sucking insect;  
 KW Diptera.

XX Synthetic.

XX Key Location/Qualifiers  
 FT Misc-difference 2 /note= "optionally absent"

XX WO200062792-A2.

XX 26-OCT-2000.

XX 04-APR-2000; 2000WO-US08879.

XX 21-APR-1999; 99US-0295924.

XX (INSECT) INSECT BIOTECHNOLOGY INC.  
 PA (UNIV) UNIV FLORIDA.

XX Bennett J, Brandt A, Borovsky D;

XX WPI, 2001-006951/01.

PT New polypeptide is useful for preventing, reducing and eliminating  
 PT infestation of area by pests e.g. flesh flies or mosquito larvae,

PT optionally in combination with e.g. repellent, attractant, acaricide,  
 PT fungicide or herbicide -

XX Example; Page 49; 87pp; English.

CC The present sequence represents an analogue of a peptide hormone named  
 CC trypsin modulating oostatic factor (TMOF). Mosquitoes regulate trypsin  
 CC and trypsin-like enzymes with this hormone to conserve metabolic  
 CC energy. Other insect pests use it for the same purpose. The peptide  
 CC functions as a pesticidal compound, which inhibits digestion in pests by  
 CC inhibiting synthesis of pests digestive enzymes. The pesticidal  
 CC compound is useful for preventing, reducing or eliminating infestation  
 CC of geographical areas by an insect population such as flesh flies, fleas,  
 CC sand flies, house flies and dog flies. The pesticidal compound is applied  
 CC to pest inhabited loci of the geographical area such as the body of  
 CC water inhabited by mosquito larvae, or insects such as coleopterans,  
 CC lepidopterans, dipterans or blood-sucking insects of order Diptera,  
 CC suborder Nematocera, family Colicidae or subfamily Culicinae,  
 CC Corethrinae, Ceratopogonidae and Simuliidae.

SQ Sequence 23 AA;

Query Match 0.7%; Score 9; DB 22; Length 23;  
 Best Local Similarity 100.0%; Pred. No. 1.4;  
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 PAPP PPPP 226  
 Db 4 PAPP PPPP 12

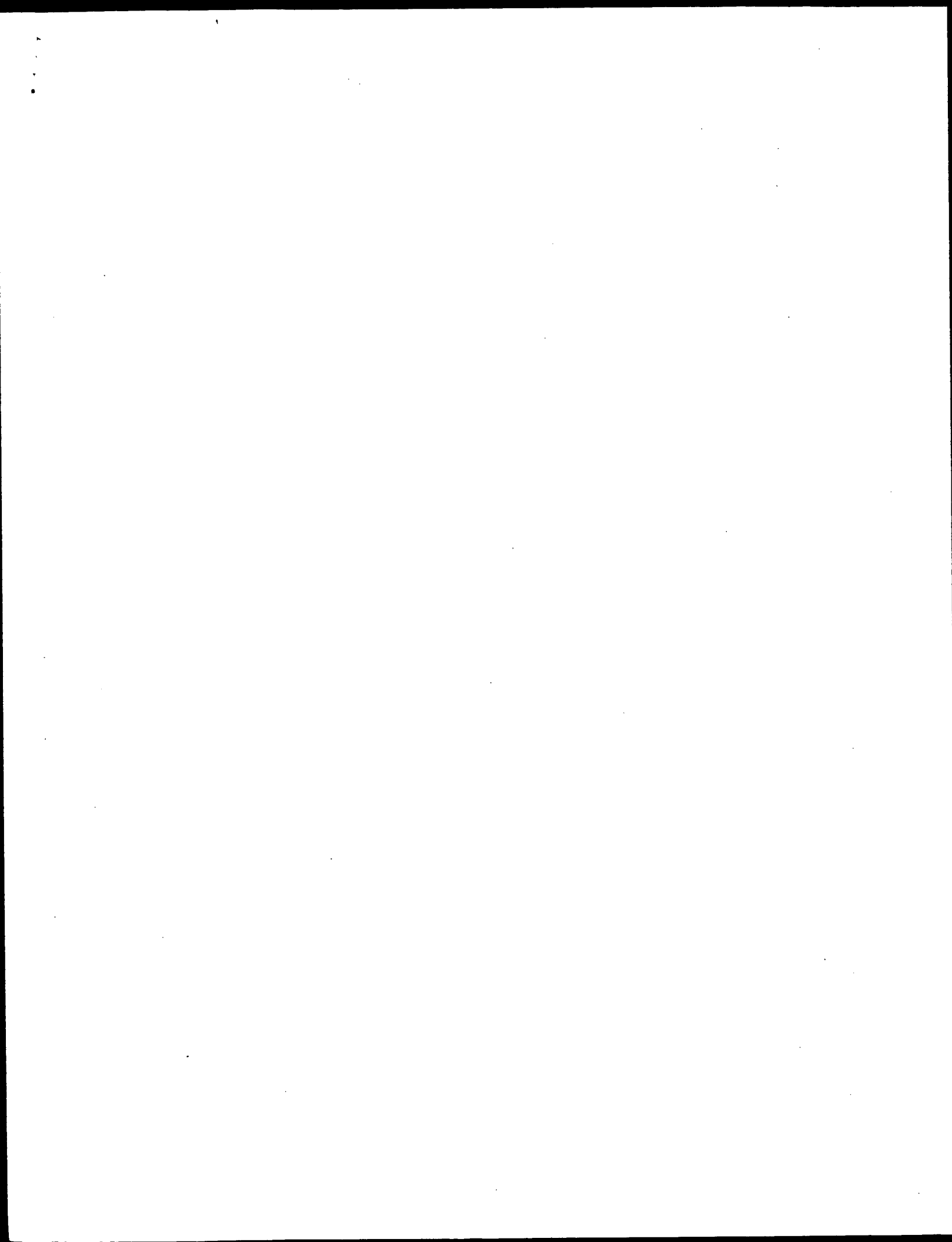
Search completed: July 19, 2001, 16:07:46  
 Job time: 206 sec

Fri Jul 20 10:09:29 2001

us-09-501-171a-4.oli.rag

---

Page 11





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 19, 2001, 16:05:20 ; Search time 15.55 Seconds  
(without alignments)  
1586,957 Million cell updates/sec

Title: US-09-501-171A-4

Perfect score: 1225  
Sequence: 1 MFARKPPGAPLGMVPDQ.....PYSELNETHSHYSPDSW 1225

Scoring table: OLIGO  
Gapop 60.0 , Gapext 60.0

Searched: 193259 seqs, 20144635 residues

Word size : 0

Total number of hits satisfying chosen parameters: 193259

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCITUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/Backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	9	0.7	15	4	US-08-602-999A-423
2	9	0.7	210	1	US-08-234-783-2
3	9	0.7	210	1	US-08-456-907-2
4	9	0.7	210	5	PCT-US95-05523-2
5	9	0.7	429	1	US-08-234-783-4
6	9	0.7	429	1	US-08-456-907-4
7	9	0.7	429	5	PCT-US95-05523-4
8	9	0.7	449	1	US-08-102-942A-4
9	9	0.7	449	1	US-08-102-942A-6
10	9	0.7	539	3	US-08-906-360-1
11	9	0.7	928	1	US-08-204-328-1
12	9	0.7	928	2	US-08-959-638-8
13	9	0.7	928	2	US-08-482-627-5
14	9	0.7	928	3	US-08-801-092-4
15	9	0.7	928	4	US-08-328-673A-8
16	9	0.7	928	5	PCT-US94-10357-2
17	9	0.7	928	5	PCT-US94-10357-3
18	9	0.7	1323	5	PCT-US94-10357-3
19	9	0.7	11	3	US-08-026-138E-4
20	8	0.7	11	3	US-08-105-904B-9
21	8	0.7	11	4	US-08-105-904B-21
22	8	0.7	11	4	US-08-114-877A-9
23	8	0.7	11	4	US-08-114-877A-14
24	8	0.7	15	4	US-08-602-999A-348
25	8	0.7	15	4	US-08-602-999A-352
26	8	0.7	15	4	US-08-602-999A-381
27	8	0.7	18	4	US-08-602-999A-422

28	8	0.7	19	4	US-08-602-999A-408	Sequence 408, App
29	8	0.7	21	3	US-08-630-916A-108	Sequence 108, App
30	8	0.7	25	1	US-08-240-514-56	Sequence 56, App
31	8	0.7	25	2	US-08-612-302A-56	Sequence 56, App
32	8	0.7	223	1	US-08-167-035-43	Sequence 43, App
33	8	0.7	223	1	US-08-208-887A-43	Sequence 43, App
34	8	0.7	223	2	US-08-359-005-43	Sequence 43, App
35	8	0.7	421	4	US-09-093-448-4	Sequence 4, App
36	8	0.7	445	2	US-08-900-148-2	Sequence 2, App
37	8	0.7	472	2	US-08-976-255-17	Sequence 17, App
38	8	0.7	478	2	US-09-040-799-3	Sequence 3, App
39	8	0.7	478	4	US-09-093-448-1	Sequence 1, App
40	8	0.7	478	4	US-09-093-448-2	Sequence 2, App
41	8	0.7	478	4	US-09-093-448-3	Sequence 3, App
42	8	0.7	480	4	US-09-108-020-6	Sequence 6, App
43	8	0.7	549	1	US-08-354-240A-2	Sequence 2, App
44	8	0.7	550	1	US-08-354-240A-4	Sequence 4, App
45	8	0.7	550	1	US-08-354-240A-6	Sequence 6, App

## ALIGNMENTS

RESULT 1  
US-08-602-999A-423  
; Sequence 423, Application US/08602999A  
; Patent No. 6184205  
; GENERAL INFORMATION:  
; APPLICANT: SPARKS, Andrew B.  
; APPLICANT: KAY, Brian K.  
; APPLICANT: THORN, Judith M.  
; APPLICANT: QUILTER, Lawrence A.  
; APPLICANT: DER, Channing J.  
; APPLICANT: FOWLER, Dana M.  
; APPLICANT: RIDER, James E.  
; TITLE OF INVENTION: SH3 BINDING PEPTIDES AND METHODS OF  
; TITLE OF INVENTION: ISOLATING AND USING SAME  
; NUMBER OF SEQUENCES: 467  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentln Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/602,999A  
; FILING DATE: 16-FEB-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mistrock, S. Leslie  
; REGISTRATION NUMBER: 18,872  
; REFERENCE/DOCKET NUMBER: 1101-202  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-9741/8864  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 423:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 15 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; MOLECULE TYPE: peptide  
; US-08-602-999A-423  
; Query Match 0.7%; Score 9; DB 4; Length 15;  
; Best Local Similarity 100.0%; Pred. No. 0.43;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 218 PAPPPEPP 226  
|||||  
Db 3 PAPPPEPP 11

## RESULT 2

US-08-234-783-2  
; Sequence 2, Application US/08234783  
; Patent No. 5622835  
; GENERAL INFORMATION:  
; APPLICANT: Herlyn, Meenhard  
; APPLICANT: Morris, Jennifer  
; APPLICANT: Rauscher III, Frank J.  
; APPLICANT: Rodeck, Ulrich  
; TITLE OF INVENTION: Wt1 Monoclonal Antibodies and Methods of  
; TITLE OF INVENTION: Use Therefor  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Howson and Howson  
; STREET: Spring House Corporate Cntr, PO Box 457  
; CITY: Spring House  
; STATE: Pennsylvania  
; COUNTRY: USA  
; ZIP: 19477  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/234,783  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bak, Mary E.  
; REGISTRATION NUMBER: 31,215  
; REFERENCE/DOCKET NUMBER: MST48USA  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 215-540-9200  
; TELEFAX: 215-540-5818  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 210 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-234-783-2

Query Match 0.7%; Score 9; DB 1; Length 210;  
Best Local Similarity 100.0%; Pred. No. 4.4;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 PAPPPEPP 226  
|||||  
Db 69 PAPPPEPP 77

## RESULT 3

US-08-456-907-2  
; Sequence 2, Application US/08456907  
; Patent No. 5633142  
; GENERAL INFORMATION:  
; APPLICANT: Herlyn, Meenhard  
; APPLICANT: Morris, Jennifer  
; APPLICANT: Rauscher III, Frank J.  
; APPLICANT: Rodeck, Ulrich  
; TITLE OF INVENTION: Wt1 Monoclonal Antibodies and Methods of  
; TITLE OF INVENTION: Use Therefor  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:

ADDRESSEE: Howson and Howson  
STREET: Spring House Corporate Cntr, PO Box 457  
CITY: Spring House  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19477

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/456,907  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/234,783  
FILING DATE: 28-APR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Bak, Mary E.  
REGISTRATION NUMBER: 31,215  
REFERENCE/DOCKET NUMBER: MST48AUSA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-540-9200  
TELEFAX: 215-540-5818  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 210 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-456-907-2

Query Match 0.7%; Score 9; DB 1; Length 210;  
Best Local Similarity 100.0%; Pred. No. 4.4;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 PAPPPEPP 226  
|||||  
Db 69 PAPPPEPP 77

## RESULT 4

PCT-US95-05523-2  
; Sequence 2, Application PC/TUS9505523  
; GENERAL INFORMATION:  
; APPLICANT: The Wistar Institute of Anatomy and Biology  
; TITLE OF INVENTION: Wt1 Monoclonal Antibodies and  
; TITLE OF INVENTION: Methods of Use Therefor  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Howson and Howson  
; STREET: Spring House Corporate Cntr, PO Box 457  
; CITY: Spring House  
; STATE: Pennsylvania  
; COUNTRY: USA  
; ZIP: 19477  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/05523  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/234,783  
; FILING DATE: 28-APR-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bak, Mary E.  
; REGISTRATION NUMBER: 31,215

REFERENCE/DOCKET NUMBER: MST48PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-540-9200  
TELEFAX: 215-540-5818  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 210 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US95-05523-2

Query Match 0.7%; Score 9; DB 5; Length 210;  
Best Local Similarity 100.0%; Pred. No. 4.4;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 218 PAPPPEPP 226  
|||||  
DB 69 PAPPPEPP 77

RESULT 5  
US-08-234-783-4  
Sequence 4, Application US/08234783  
Patent No. 5622835  
GENERAL INFORMATION:  
APPLICANT: Herlyn, Meenhard  
APPLICANT: Morris, Jennifer  
APPLICANT: Rauscher III, Frank J.  
APPLICANT: Rodeck, Ulrich  
TITLE OF INVENTION: Wt1 Monoclonal Antibodies and Methods of  
NUMBER OF SEQUENCES: 7  
TITLE OF INVENTION: Use Therefor  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Howson and Howson  
STREET: Spring House Corporate Cntr, PO Box 457  
CITY: Spring House  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19477  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/234,783  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Bak, Mary E.  
REGISTRATION NUMBER: 31,215  
REFERENCE/DOCKET NUMBER: MST48USA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-540-9200  
TELEFAX: 215-540-5818  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 429 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-234-783-4

Query Match 0.7%; Score 9; DB 1; Length 429;  
Best Local Similarity 100.0%; Pred. No. 8.2;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 218 PAPPPEPP 226  
|||||  
DB 58 PAPPPEPP 66

RESULT 6  
US-08-456-907-4  
Sequence 4, Application US/08456907  
Patent No. 5633142  
GENERAL INFORMATION:  
APPLICANT: Herlyn, Meenhard  
APPLICANT: Morris, Jennifer  
APPLICANT: Rauscher III, Frank J.  
APPLICANT: Rodeck, Ulrich  
TITLE OF INVENTION: Wt1 Monoclonal Antibodies and Methods of  
NUMBER OF SEQUENCES: 7  
TITLE OF INVENTION: Use Therefor  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Howson and Howson  
STREET: Spring House Corporate Cntr, PO Box 457  
CITY: Spring House  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19477  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/456,907  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/234,783  
FILING DATE: 28-APR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Bak, Mary E.  
REGISTRATION NUMBER: 31,215  
REFERENCE/DOCKET NUMBER: MST48USA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 215-540-9200  
TELEFAX: 215-540-5818  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 429 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-456-907-4

Query Match 0.7%; Score 9; DB 1; Length 429;  
Best Local Similarity 100.0%; Pred. No. 8.2;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 218 PAPPPEPP 226  
|||||  
DB 58 PAPPPEPP 66

RESULT 7  
PCT-US95-05523-4  
Sequence 4, Application PC/TUS9505523  
GENERAL INFORMATION:  
APPLICANT: The Wistar Institute of Anatomy and Biology  
TITLE OF INVENTION: Wt1 Monoclonal Antibodies and  
TITLE OF INVENTION: Methods of Use Therefor  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Howson and Howson  
STREET: Spring House Corporate Cntr, PO Box 457  
CITY: Spring House  
STATE: Pennsylvania  
COUNTRY: USA  
ZIP: 19477

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/05523
; FILING DATE:
; CLASSIFICATION:
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 08/234,783
; FILING DATE: 28-APR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Bak, Mary E.
; REGISTRATION NUMBER: 31,215
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-540-9200
; TELEFAX: 215-540-5818
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 429 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US95-05523-4

Query Match          0.7%; Score 9; DB 5; Length 429;
Best Local Similarity 100.0%; Pred. No. 8.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 PAPPPEPP 226
   |||||||
Db 58 PAPPPEPP 66

RESULT 8
US-08-102-942A-4
; Sequence 4, Application US/08102942A
; Patent No. 5726288
; GENERAL INFORMATION:
; APPLICANT: Call, Katherine M.
; APPLICANT: Glaser, Thomas M.
; APPLICANT: Ito, Caryn Y.
; APPLICANT: Buckler, Alan J.
; APPLICANT: Pelletier, Jerry
; APPLICANT: Haber, Daniel A.
; APPLICANT: Rose, Elise A.
; APPLICANT: Housman, David E.
; APPLICANT: Bruening, Wendy
; APPLICANT: Darveau, Andre
; TITLE OF INVENTION: Localization and Characterization of the
; TITLE OF INVENTION: Wilms' Tumor Gene
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Millitia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: U.S.
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/102,942A
; FILING DATE: 02-AUG-1993
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; 
```

```

; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: MIT-5194A2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 449 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-102-942A-4

Query Match          0.7%; Score 9; DB 1; Length 449;
Best Local Similarity 100.0%; Pred. No. 8.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 PAPPPEPP 226
   |||||||
Db 59 PAPPPEPP 67

RESULT 9
US-08-102-942A-6
; Sequence 6, Application US/08102942A
; Patent No. 5726288
; GENERAL INFORMATION:
; APPLICANT: Call, Katherine M.
; APPLICANT: Glaser, Thomas M.
; APPLICANT: Ito, Caryn Y.
; APPLICANT: Buckler, Alan J.
; APPLICANT: Pelletier, Jerry
; APPLICANT: Haber, Daniel A.
; APPLICANT: Rose, Elise A.
; APPLICANT: Housman, David E.
; APPLICANT: Bruening, Wendy
; APPLICANT: Darveau, Andre
; TITLE OF INVENTION: Localization and Characterization of the
; TITLE OF INVENTION: Wilms' Tumor Gene
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Millitia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: U.S.
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/102,942A
; FILING DATE: 02-AUG-1993
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-861-6240
; TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 449 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-102-942A-6

```

Query Match 0.7%; Score 9; DB 1; Length 449;  
Best Local Similarity 100.0%; Pred. No. 8.6;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 PAPPPEPP 226  
|||||  
DB 58 PAPPPEPP 66

## RESULT 10

US-08-906-360-1  
; Sequence 1, Application US/08906360  
; Patent No. 6013441  
; GENERAL INFORMATION:  
; APPLICANT: Meruelo, Daniel  
; APPLICANT: Pampero, Christine  
; TITLE OF INVENTION: MAMMALIAN HUMAN FXI-T1  
; FILE REFERENCE: 8105-010  
; CURRENT APPLICATION NUMBER: US/08/906,360  
; CURRENT FILING DATE: 1997-08-05  
; EARLIER APPLICATION NUMBER: 60/023,173  
; EARLIER FILING DATE: 1996-08-05  
; NUMBER OF SEQ ID NOS: 3  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1  
; LENGTH: 539  
; TYPE: PRT  
; ORGANISM: mammalian  
US-08-906-360-1

Query Match 0.7%; Score 9; DB 3; Length 539;  
Best Local Similarity 100.0%; Pred. No. 10;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 PAPPPEPP 226  
|||||  
DB 146 PAPPPEPP 154

## RESULT 11

US-08-204-329-1  
; Sequence 1, Application US/08204329  
; Patent No. 5710255  
; GENERAL INFORMATION:  
; APPLICANT: SHEPARD, H. M.  
; APPLICANT: MEN, SHU F.  
; TITLE OF INVENTION: CHARACTERIZATION OF A NOVEL ANTI-P110RB  
; TITLE OF INVENTION: MONOCLONAL ANTIBODY  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: TOWNSEND & TOWNSEND & CREW LLP  
; STREET: TWO EMBARCADERO CENTER, 8TH FLOOR  
; CITY: SAN FRANCISCO  
; STATE: CALIFORNIA  
; COUNTRY: U.S.A.  
; ZIP: 94111  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/204,329  
; FILING DATE: 15-AUG-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US92/05866  
; FILING DATE: 14-JUL-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: RENEE A. FITTS  
; REGISTRATION NUMBER: 35,136

REFERENCE/DOCKET NUMBER: 16930-00040005

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 326-2400

TELEFAX: (415) 326-2422

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 928 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-204-329-1

Query Match 0.7%; Score 9; DB 1; Length 928;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 PAPPPEPP 226  
|||||  
DB 21 PAPPPEPP 29

## RESULT 12

US-08-959-638-8  
; Sequence 8, Application US/08959638  
; Patent No. 5932210  
; GENERAL INFORMATION:  
; APPLICANT: Gregory, Richard J.  
; APPLICANT: Mills, Ken N.  
; APPLICANT: Maneval, Daniel C.  
; TITLE OF INVENTION: Recombinant Adenoviral Vector and  
; TITLE OF INVENTION: Methods of Use  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Campbell and Flores  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: USA  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/959,638  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/328,673  
; FILING DATE: 25-OCT-1994  
; APPLICATION NUMBER: US/08/233,777  
; FILING DATE: 19-MAY-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/142,669  
; FILING DATE: 25-OCT-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Campbell, Cathryn A.  
; REGISTRATION NUMBER: 31,815  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 928 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-959-638-8

Query Match 0.7%; Score 9; DB 2; Length 928;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 PAPPPEPP 226  
DB 21 PAPPPEPP 29

RESULT 13

US-08-482-627-5  
; Sequence 5, Application US/08482627  
; Patent No. 5998134  
; GENERAL INFORMATION:  
; APPLICANT: Lee, Men-Hwa  
; APPLICANT: Lee, Eva Y-H.P  
; TITLE OF INVENTION: Retinoblastoma Gene - Cancer Suppressor  
; TITLE OF INVENTION: and Regulator  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Campbell and Flores  
; STREET: 4370 La Jolla Village Drive, Suite 700  
; CITY: San Diego  
; STATE: California  
; COUNTRY: United States  
; ZIP: 92122  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/482,627  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/951,947  
; FILING DATE: 28-SEP-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Campbell, Cathryn A.  
; REGISTRATION NUMBER: 31,815  
; TELEPHONE: (619) 535-9001  
; TELEFAX: (619) 535-8949  
; REFERENCE/DOCKET NUMBER: P-UC 1707  
; TELECOMMUNICATION INFORMATION:  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 928 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-482-627-5

Query Match 0.7%; Score 9; DB 2; Length 928;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 PAPPPEPP 226  
DB 21 PAPPPEPP 29

RESULT 14

US-08-801-092-4  
; Sequence 4, Application US/08801092  
; Patent No. 6074850  
; GENERAL INFORMATION:  
; APPLICANT: Antelman, Douglas  
; APPLICANT: Gregory, Richard J.  
; APPLICANT: Wils, Kenneth N.  
; TITLE OF INVENTION: Tissue Specific Expression of

; TITLE OF INVENTION: Retinoblastoma Protein  
; NUMBER OF SEQUENCES: 46  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP  
; STREET: Two Embarcadero Center, 8th Floor  
; CITY: San Francisco  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94111

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/801,092  
; FILING DATE: 14-FEB-1997  
; CLASSIFICATION: 514  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/751,517  
; FILING DATE: 15-NOV-1996  
; CLASSIFICATION: 514  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Flits, Renee A.  
; REGISTRATION NUMBER: 35,136  
; REFERENCE/DOCKET NUMBER: 016930-001020  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 415-576-0200  
; TELEFAX: 703-576-0300  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 928 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: peptide  
US-08-801-092-4

Query Match 0.7%; Score 9; DB 3; Length 928;  
Best Local Similarity 100.0%; Pred. No. 16;  
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 218 PAPPPEPP 226  
DB 21 PAPPPEPP 29

RESULT 15

US-08-328-673A-8  
; Sequence 8, Application US/08328673A  
; Patent No. 6210939  
; GENERAL INFORMATION:  
; APPLICANT: Gregory, Richard J.  
; APPLICANT: Wills, Ken N.  
; APPLICANT: Maneval, Daniel C.

; TITLE OF INVENTION: Recombinant Adenoviral Vector and  
; Methods of Use

; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Townsend and Townsend and Crew LLP  
; STREET: Two Embarcadero Center, Eighth Floor  
; CITY: San Francisco  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94111-3834  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/328,673A

```

: FILING DATE: 25-Oct-1994
: CLASSIFICATION: <Unknown>
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/142,669
: FILING DATE: 25-Oct-1993
: APPLICATION NUMBER: US 08/233,669
: FILING DATE: 26-Apr-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Smith, Timothy S.
: REGISTRATION NUMBER: 35,367
: REFERENCE/DOCKET NUMBER: 016930-000920US
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (415) 576-0200
: TELEFAX: (415) 576-0300
: INFORMATION FOR SEQ ID NO: 8:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 928 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-08-328-673A-8

```

```

Query Match          0.7%; Score 9; DB 4; Length 928;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 218 PAPPPIPP 226
   |||||
Db 21 PAPPPIPP 29

```

Search completed: July 19, 2001, 16:08:05  
 Job time: 165 sec

